AD_____

GRANT NUMBER DAMD17-94-J-4439

TITLE: Structural Studies of the PU.1 Transcription Factor

PRINCIPAL INVESTIGATOR: Kathryn R. Ely, Ph.D.

CONTRACTING ORGANIZATION: The Burnham Institute
La Jolla, CA 92037

REPORT DATE: October 1997

TYPE OF REPORT: Final

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1997	Final (1 Sep 94 - 31 Aug 97)

4. TITLE AND SUBTITLE
Structural Studies of the PU.1 Transcription Factor

5. FUNDING NUMBERS
DAMD17-94-J-4439

6. AUTHOR(S)
Kathryn R. Ely, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Burnham Institute
La Jolla, CA 92037

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

Ets transcription factors play a role in development and are implicated in some malignant processes. Recently, members of this large gene family have been identified in normal gene expression in mammary cells and also in breast cancer cell lines. In these studies, the crystal structure of the DNA-binding domain of the PU.1 ets protein complexed to DNA has been determined at 2.1 Å resolution. The DNA binding domain is a conserved region that binds the core sequence 5'-GGAA/T-3'. The PU.1 domain binds DNA using a loop-helix-loop motif involving conserved amino acids and bases. In this project we have also used nuclear magnetic resonance (NMR) to determine the unbound structure of the domain in solution. The two structures were correlated to understand the process of DNA recognition by ets proteins.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Transcription Factor, Multidimensional NMR, Protein-DNA Complex, X-Ray Crystallography, DNA-Binding Domain, Breast Cancer	33
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

N/A Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Kathryn R. Ely 9/29/97

PI - Signature Date

TABLE OF CONTENTS

Annual Report -- Grant DAMD17-94-J-4439

INTRODUCTION

Significance. The ets gene family, a recently discovered family of regulatory proteins, includes more than 45 members in a variety of organisms from Drosophila to humans (Wasylyk et al., 1993; Moreau-Gachelin, 1994). These molecules play a role in normal development and have been implicated in malignant processes such as leukemia or breast cancer. Enthusiasm is quite strong for the study of ets proteins in cancer research (Hromas and Klemsz, 1994) because the family is large and composed of individual members that are distinct and function as regulatory proteins in a variety of cell types. With respect to breast cancer, ets-related proteins have been identified in normal mammary cell-specific gene expression (Welte et al, 1994) as well as in breast cancer cell lines (Trimble et al., 1993; Slamon et al., 1989; Scott et al, 1989). An interesting association of ets proteins with malignant transformation has been suggested in the observations that the phosphoprotein osteopontin is regulated by ets-related proteins (Denhardt and Guo, 1993; Guo et al., 1995). Expression of this protein is also responsive to hormones such as estrogen and progesterone. The expression level of osteopontin is significantly elevated in transformed cells and is related to the metastatic potential of the tumor cells (Guo et al., 1995; Brown et al., 1994; Gardner et al., 1994). These facts suggest a possible mechanism whereby ets-related proteins may be implicated in the development and/or metastatic spread of breast tumors.

Background. The PU.1 transcription factor is an ets protein expressed in hematopoietic cells (Klemsz et al., 1990). The ets proteins share a conserved domain of around 85 amino acids which binds as a monomer to the DNA sequence: 5'-C/AGGAA/T-3'. Within the ets family, the PU.1 sequence is the most divergent from ets-1 and yet there is 40% sequence homology in the DNA-binding domain ofthese two proteins. We have selected the PU.1 ets DNA-binding domain for structural studies using both crystallography and nuclear magnetic resonance (NMR) to derive the data. The crystallographic analyses were focused on the structure of the protein-DNA complex and the NMR work highlighted the domain in solution, to evaluate dynamic aspects of the structure.

BODY -- FINAL REPORT

Task 1: Large scale purification of the PU.1 DNA-binding domain

The DNA-binding domain of PU.1 was cloned in the pET11 expression vector by polymerase chain reaction amplification of the DNA-binding

domain from the full length cDNA as described previously (Klemsz et al., 1990). For bacterial expression, pET plasmid constructs were used to transform E. coli BL21(DE3)pLysS cells. The protein was expressed in large scale cultures and purified by ion-exchange chromatography and gel filtration. This protein, representing residues 160-272, was used for crystallization in complex with DNA. Attempts to co-crystallize a shorter fragment with DNA were unsuccessful. However, for NMR analysis another protein fragment was generated. This short fragment of 93 residues was highly soluble and amenable for the high protein concentrations required for the NMR experiments outlined in Task 3. The conclusion from the protein purification work is that length of the recombinant fragment is extremely critical for successful crystallization of the protein-DNA complex (published observations in Pio et al., 1995) and also to produce a tight globular domain for clear resonance dispersion in NMR analyses.

Task 2: Synthesis of DNA oligonucleotides

DNA oligonucleotides of various lengths were screened for binding in complex with the PU.1 domain. The quality of the oligonucleotides was critical for successful co-crystallization. Oligonucleotides were synthesized on a 10 μM scale using phosphoramidite chemistry and an automated DNA synthesizer. Oligos were purified by reverse phase HPLC at 56°C in an acetonitrile gradient. After removing the acetonitrile by dialysis against triethylammonium bicarbonate buffer, the oligonucleotides were desalted in ethanol on phosphocellulose resin and lyophilized. For the series of oligonucleotides, each one differed in length and contained the core sequence GGAA which is the recognition sequence for PU.1. Oligos were designed to provide both blunt-ended duplex DNA fragments and fragments that had unpaired T or A bases at the termini. The latter were tested because they have the potential for end-to-end stacking in the crystal lattice. Ultimately a sixteen base-pair oligonucleotide with the sequence 5'-AAAAAGGGGAAGTGGG-3' and the complementary strand were selected and synthesized on the large scale, purified and annealed together into duplex DNA. This oligo promoted the formation of crystals of the complex (results published in Pio et al., 1995). It was evident in the electron density map of the complex that the DNA fragments formed long extended fiber-like elements in the crystal lattice by end-to-end stacking between adjacent oligonucleotides, and that this was a major interaction for nucleation of crystal growth.

Task 3: Determination of the solution structure of the PU.1 domain by NMR

Work with a short fragment of the domain, produced to optimize the NMR studies, permitted the production of doubly-labeled (^{13}C and ^{15}N) sample for the unambiguous assignment of all resonances in the 93 residue domain. In heteronuclear experiments all atoms for the polypeptide backbone have been assigned: ^{15}NH, $^{13}C\alpha$, and ^{13}CO. Extension to the side chain resonances is beyond 70% complete at this stage. More than 850 nuclear Overhauser effects (NOE) are observed. Each NOE represents a uniquely identified interaction between two protons and depends principally on the distance between two nuclei. This distance analysis gives information on the three-dimensional shape of the molecule in its folded conformation. For the PU.1 domain, an average of nine NOEs per residue were observed. Based on these data, all secondary structure elements have been defined. At this point, we do not see large differences between the free PU.1 domain in solution and the domain bound to DNA in the crystal structure. However, by NMR, structural assignments can be made for residues at the amino- and carboxyl-terminal regions that do not contact DNA. These regions were disordered or flexible in the crystal and were not seen in the electron density maps. All unambigous assignments are listed in Table 1.

To derive assignments for the polypeptide backbone and to define secondary structural elements of the domain, the backbone amide resonances were analyzed in Heteronuclear Single Quantum Coherence (HSQC) experiments. In the folded protein, some proton are protected and shifts result from protecting the protons from the exchange with solvent. Thus, using experiments designed to measure the chemical shifts in a protein, it is possible to deduce information about the secondary structural features of the molecules, correlated with accessibility to the protein surface and surrounding solvent. Figure 1 presents the NOE and chemical shift data for the PU.1 domain. In addition to the expected protection of helical and β-strand elements, partial protection of residues between the first and second β-strands was observed, as well as residues at the carboxyl terminus. Interestingly, amide protons in helix α2 were not protected. This observation may be due to increased mobility in the time regime (milliseconds to hours) measured by the proton-deuterium exchange. However, tryptophan 215 in helix 3 does contact the DNA backbone. Other structural elements that were seen to contact the DNA backbone in the crystal structure of the protein-DNA complex were observed to be mobile by hydrogen-deuterium exchange; for example, the

Table I

Res.	15N	$13C_\alpha$	$13C_\beta$	13C=O	1HN	$1H_\alpha$	$1H_\beta$	Other
S169								
K169	122.4	55.8	33.3	176.1	7.79			
K170	127.4	56.2	32.9	176.2	8.36			
K171	128.0	56.0	32.9	175.8	8.39			
I172	127.2	60.4	38.9	175.0	8.15			
R173	128.6	54.5	32.3		8.28	4.36		
L174		58.1	39.6	177.7				
Y175	114.1	60.0	36.5	174.6	7.10	4.02		
Q176	125.7	57.1	28.7	176.5	5.89	3.69	1.99,1.85	2.27
F177	124.9	60.3	39.2	176.6	7.79	4.32	3.22,2.48	
L178	120.4	57.2	40.8	177.1	7.68	3.47		
L179	123.8	57.4	41.1	177.8	7.56	3.68	1.96,0.87	
D180	122.7	57.2	39.7	179.9	8.46	4.31	2.75,2.55	
L181	125.8	58.0	41.5	177.1	7.55			
L182	123.8	57.6	41.8	181.1	7.58			
R183	122.4	59.2	30.1	178.1	8.77	4.39	1.98	
S184	117.4	59.3	63.9	175.5	8.06	4.45	4.00	
G185	114.9	46.2		173.8	7.86	4.02		
D186	124.1	53.7	42.5	175.0	8.12			
M187	121.5	55.8	30.0	177.2	8.76	4.36	2.25,2.0	2.44
K188	122.4	58.8	32.1	176.5	7.94		1.93	
D189	118.2	54.2	40.0	176.9	8.34	3.96	2.64	
S190	118.5	60.8	65.6	171.8	8.57	4.02	3.92	
I191	126.9	59.3	40.5	171.3	8.23	4.11		
W192	124.9	56.0	30.1	174.9	7.85	4.87	3.22	6.92,10.07 7.29,7.09 6.84,7.54
W193	122.4	57.2	30.9	177.4	8.96	5.04	3.39,3.13	7.53,9.97 7.21,7.06 7.18,8.18
V194	128.9	64.3	31.5	176.0	9.45	4.35	1.86	1.05
D195	122.4	53.0	41.4		7.81		3.05,2.50	
K196	131.5	59.6	31.6	177.8	9.06	4.80		
D197	121.8	56.6	40.2	177.4	8.09	4.28	2.71,2.65	
K198	120.2	55.5	33.2		7.36	4.37	1.98,1.68	1.49,1.37
G199	112.1	47.4		174.3	8.05	4.32		
T200	117.4	60.3	69.6	174.6	8.37	5.49	4.08	0.98
F201	129.3	55.7	43.1	170.0	9.39	5.49	2.67,2.55	
Q202	124.0	52.8	34.0	172.4	8.65	5.10	1.57,1.30	2.06
F203	121.3	54.2	41.6	183.2	7.42	5.00	3.0,2.81	
S204	116.3	56.6	65.0	176.5	7.56	4.65	3.95	
S205	127.7	58.6	68.9	177.8	8.95	4.70	3.98	
K206	129.6	58.2	32.4	177.8	9.01	4.80	1.82	
H207	117.9	54.9	30.0	175.6	7.60	5.29	2.98	7.37,8.41
K208	124.6	59.5	30.6	177.3	7.42	4.17		
E209	123.5	58.7	27.8	178.8	8.46	4.41	2.09,2.01	2.44
A210	126.8	54.7	17.6	181.0	7.83	4.22	1.57	
L211	124.3	58.1	41.7	177.8	7.57	4.11	1.79	1.62
A212	125.7	55.0	17.5	179.0	7.80	4.42	1.42	
H213	120.4	58.3	28.7	177.4	8.63	4.48	3.48,3.42	7.44,8.48
R214	123.5	58.8	29.3	178.3	7.86	4.14	2.39,2,26	2.05
W215	126.3	58.1	29.7		8.28	4.36	3.29	6.92,10.58 7.04,6.48 6.18,7.49
G216	108.5	47.4		175.7	8.31	3.81,3.14		
I217	123.8	63.9	37.4	178.7	7.54	3.71		
Q218	124.9	57.6	27.7	177.1	7.69	3.90	1.97,1.76	2.45,2.33
K219	121.3	56.3	31.4		7.77	3.70		

G220	110.2	45.6		174.0	7.51	4.04,3.71				
N221	122.4	52.9	38.4	177.3	8.01	4.60	2.68,2.50			
R222	124.0	57.0	29.8	176.4	8.45	4.04	1.85,1.78	1.58	3.17	
K223	122.4	55.8	33.3	176.0	7.78	4.32	1.79,1.71	1.36		
K224	125.5	56.5	32.2	175.7	8.21					
M225	128.8	54.0	30.7	182.0	8.36	4.12				
T226	117.9	58.5	71.9	183.3	6.99	4.87	4.30			
Y227	128.6	63.2	37.6	176.1	10.30	3.87	2.97.2.71	6.79,6.25		
Q228	122.1	59.4	28.0	178.7	8.97	3.82				
K229	123.2	59.6	33.1	179.3	7.56	3.82				
M230	127.9	59.4	33.7	177.2	8.13	3.91	2.04	2.54,2.29		
A231	125.4	54.7	16.8	180.3	9.12	3.72	1.10			
R232	124.0	59.2	29.5	178.2	7.61	4.00	1.96	1.79,1.66		
A233	126.8	54.8	18.1	180.8	7.41	4.12	1.53			
L234	122.7	57.7	41.3	179.9	8.25	4.11	1.76,1.61	1.39		
								0.48,0.45		
R235	124.6	59.0	29.7	177.5	8.23	4.18	1.78,1.65	1.43,1.32		
N236	123.8	55.5	37.7	177.9	7.88	4.14	2.87			
Y237	122.1	60.0	38.6	177.0	7.74	4.94	3.53,3.23			
G238	112.7	47.1		176.0	8.23	4.18				
K239	123.2	58.3	32.7		7.76	4.34	2.08,1.80	1.66,1.57		
T240	112.0	62.7	67.7	177.1	7.63	4.00	3.70			
G241	114.1	46.1		182.5	8.37	4.40,3.65				
E242	126.5	61.9	29.4	176.4	10.41	3.78				
V243	114.1	60.4	35.6	182.2	6.72	4.70	1.88	0.83,0.75		
K244	129.4	54.2	35.4	182.9	8.97	4.80	1.82	1.45,1.40	1.71	
								3.25,3.13		
K245	128.9	56.0	33.1	176.3	8.75	4.01	1.85	1.36,1.06	1.67	
V246	127.7	60.3	32.3	183.0	7.21	4.39	1.88	0.79,0.57		
K247	126.9	57.7	31.3	176.5	8.03	4.13	1.86	1.46,1.38	1.73	
K248	124.9	56.1	33.8	175.7	7.86	4.23	1.61	1.46,1.30		
K249	128.0	57.9	32.5	177.6	8.30	4.12	1.80			
L250	125.7	56.1	40.0	181.4	9.24	3.91	2.06,1.95	1.63		
T251	115.2	62.1	69.0	172.4	7.13	4.97	3.51	0.93		
Y252	129.9	56.2	43.9	180.7	8.36	4.33	3.03			
Q253	122.7	54.0	33.4	176.4	9.06	5.20	1.84,1.58	2.40,2.28		
F254	131.3	58.6	41.5	175.8	8.93	5.28	3.39,2.96			
S255	120.4	57.3	64.1	175.4	8.27	4.47	4.19			
G256	113.8	47.2		176.6	8.87	3.98,3.82				
E257	123.2	59.0	29.3	178.4	8.35	4.11	2.00	2.30,2.13		
V258	119.6	63.1	31.5	176.4	7.36	4.03	2.31	0.86,0.62		
L259	123.2	55.8	42.9	177.0	7.32	4.05	1.74	1.66		
G260	116.0	46.2			7.44	3.74				

turn in the helix-turn-helix motif and the 'wing' between β-strands 3 and 4. Amide protons in the recognition helix 3 were less protected in these studies than those of helix 1 which does not contact the DNA. The results of the chemical shift experiments correlate well with the crystal structure and suggest that no major structural refolding occurs in the domain on binding DNA.

To probe backbone dynamics, NMR relaxation techniques were performed. Because motion and flexibility can markedly influence DNA recognition by DNA-binding proteins, such measurements can provide important insight into the dynamics of the protein-DNA contacts. Individual quantitative measurements of the dynamic behavior for each of the amide proton resonances from the polyppetide backbone were made. These measurements probe a much faster time regime than the hydrogen-deuterium experiments and can provide motional information in the picosecond-nanosecond scale as well as the microsecond-millisecond time regimes. A flexible element can adopt multiple conformations and thus facilitates binding to the target molecule (in this case DNA). Analyses of these relaxation data to date are presented in Figure 2. Relaxation rates (R1 and R2) and NOE intensities indicate a higher intrinsic flexibility in the loop between helices 2 and 3. We reported that this connecting segment is actually a loop and intermediate in length relative to the counterpart in other members of the HTH family (Pio et al., 1996). In PU.1, this loop which is seven residues long, contacts DNA, and has the lowest NOE intensities and relaxation rates of the domain.

Figure 1: Structural elements for the free DNA-binding domain of PU.1, determined by NMR. The sequence of the domain is listed on the top line. Underneath the sequence, Lines 2-5 indicate the sequential NOE correlations (intensities). Lines 6-7 present chemical shift changes relative to the corresponding random coil values for Cα and Hα. Amide resonances that could be detected in the first HSQC spectrum in the series of ^{1}H-^{2}H exchange experiments are marked with open circles in Line 8, while residues that are observable after several hours are indicated by filled circles. The inferred secondary structural elements along the sequence are listed at the bottom of the figure.

Figure 2: Relaxation measurements for the PU.1 domain. NOE, R1, R2 and S^2 values are plotted for each of the residues in the domain. R1 and R2 are relaxation rates and S^2 is an order parameter that gives a measure of the rigidity of an element. Values of $S^2 > 0.85$ indicate greater conformational restriction. Note that the S^2 values were considerably low for three loops: the loop between helix 2 and 3 which is the DNA recognition helix, the loop between β-strand 3 and 4 which contacts the DNA and the loop between helix 3 and β-strand 3 (no DNA contacts). In R2, the asterisk indicates that the residue M187 has a weak crosspeak in the ^{15}N-HSQC spectrum. In S^2, two asterisks indicate two residues which are not resolved.

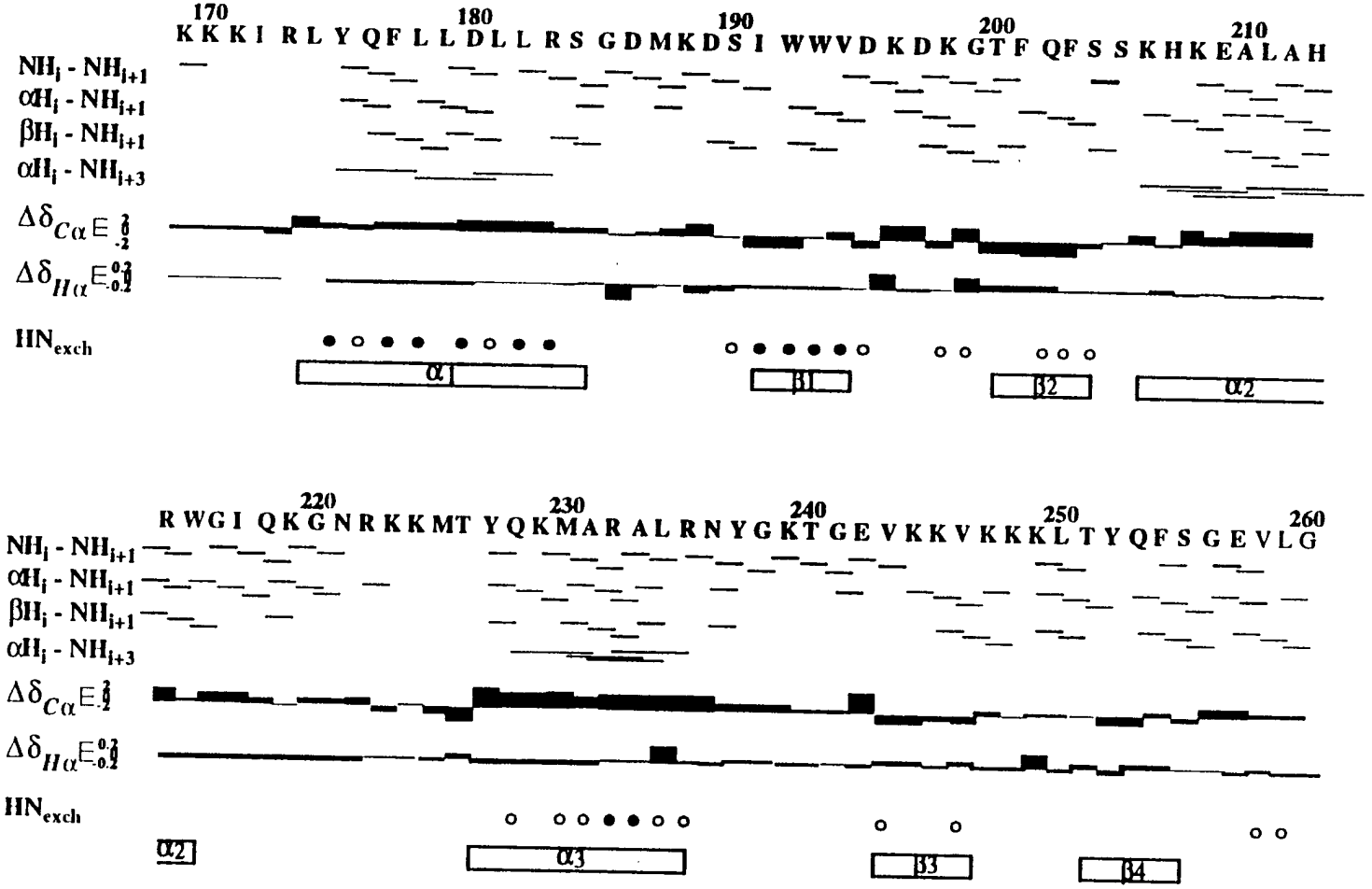
170
180
190
200
210
K K K I R L Y Q F L L D L L R S G D M K D S I W W V D K D K G T F Q F S S K H K E A L A H
NHi - NHi+1
αHi - NHi+1
βHi - NHi+1
αHi - NHi+3
ΔδCα
ΔδHα
HNexch
α1
β1
β2
α2
220
230
240
250
260
R W G I Q K G N R K K M T Y Q K M A R A L R N Y G K T G E V K K V K K K L T Y Q F S G E V L G
NHi - NHi+1
αHi - NHi+1
βHi - NHi+1
αHi - NHi+3
ΔδCα
ΔδHα
HNexch
α2
α3
β3
β4

Figure 1

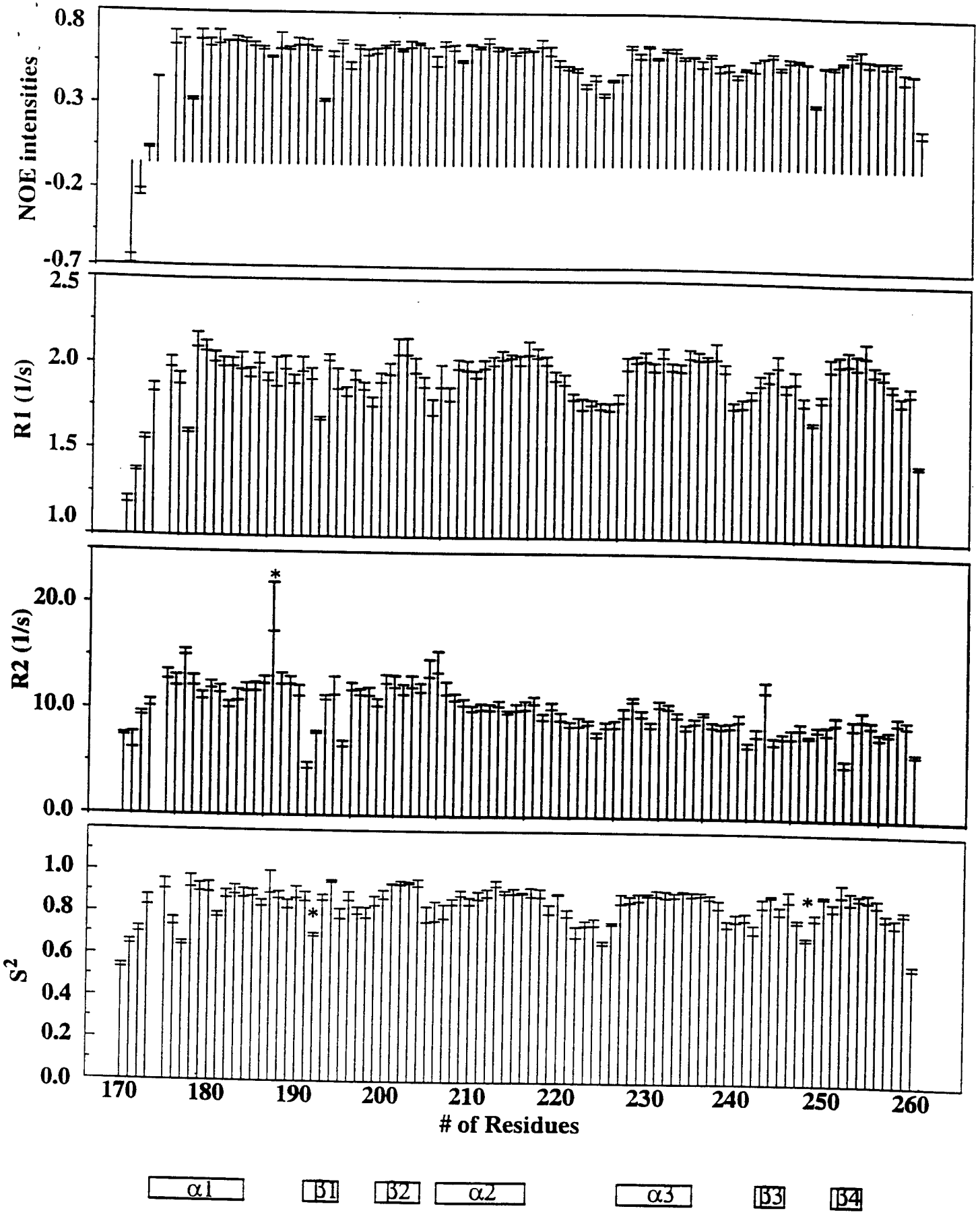
NOE intensities
0.8
0.3
-0.2
-0.7
R1 (1/s)
2.5
2.0
1.5
1.0
R2 (1/s)
20.0
10.0
0.0
S2
1.0
0.8
0.6
0.4
0.2
0.0
170
180
190
200
210
220
230
240
250
260
# of Residues
α1
β1
β2
α2
α3
β3
β4

Figure 2

Task 4: Determination of the crystal structure of the PU.1 domain complexed to DNA

The PU.1-DNA complex crystallized in the space group C2 with a= 89.1, b=101.9, c=55.6 Å and β=111.2°, with two complexes in the asymmetric unit. Crystal growth was induced from solutions containing cacodylate buffer at pH 6.5, zinc acetate and polyethyleneglycol 600 (published data in Pio et al., 1995). The crystals diffracted to 2.3 Å resolution (2.1Å at the LURE synchrotron). Four heavy atom derivatives were prepared by soaking crystals in mercury compounds and by co-crystallizing with iodinated oligonucleotides. The structure was solved by the multiple isomorphous replacement method plus anomalous scattering from the mercury compound (MIRAS). An atomic model fitted to electron density maps calculated at 2.3 Å resolution (2.1Å refined) revealed the structure of the complex and was reported (Kodandapani et al., 1996). This is the first (and only) report of a crystal structure of an ets protein.

The PU.1 domain assumes a tight globular structure (33 x 34 x 38 $Å^3$) formed by three α-helices and a four-stranded antiparallel β-sheet. The domain topology is similar to the structure of other ets family proteins fli-1 (Liang et al., 1994), murine ets-1 (Donaldson et al., 1996), and human ets-1 (Werner et al., 1995) determined in solution by NMR. The structures revealed that ets domains share a common folding pattern that is similar to α+β helix-turn-helix (HTH) DNA-binding proteins and resembles 'winged' HTH proteins such as HNF-3γ (Clark et al., 1993). The domain contacts DNA from three sites: the recognition helix (α3), the loop between β-strands 3 and 4 (a 'wing'), and the turn in the HTH motif (α2-turn-α3). This turn is longer that the equivalent in many other HTH proteins and is actually a loop. Our structure revealed a new pattern for HTH recogntion and a novel mode of DNA binding (reported in Kodandapani et al., 1996).

The DNA is bent in the complex (8°) when compared to 'canonical' B-DNA structure and is curved uniformly along the entire 16 bp length. The minor groove is slightly enlarged (~2-3 Å from the mean) in the GGAA region at the midpoint of the oligonucleotide. There was a report that a human ets-1-DNA complex was quite different where the protein contacted the DNA (Werner et al., 1995) with a kinked deformation of the DNA by 60°, however, this model was found to be in error due to a misinterpretation of the NMR data in their structure solution. The authors of that structure retracted the model (Werner et al., 1996 erratum) and reported that the ets-1-DNA complex was in fact quite similar to our PU.1-DNA structure.

Four strictly conserved residues on the surface of the domain are likely to be important for DNA-binding by all members of the ets family. Arg 232 and Arg 235 emanate from the recognition helix and contact the conserved GGAA sequence in the major groove of DNA. Lys 245 from the 'wing' contacts the phosphate backbone of the GGAA strand in the minor groove upstream from the core sequence and Lys 219 in the loop of the HTH motif forms a salt bridge with the phosphate backbone of the opposite strand downstream of the GGAA core. Substitutions of glycine at each of these four conserved sites abolished DNA binding, confirming the functional importance of these residues. These interactions were further evaluated by mutagenesis of PU.1 and comparisons of mutagenesis on other ets molecules. We reported (Pio et al., 1996) that these interactions represent the paradigm for ets recognition which is expected to be reproduced in all ets proteins.

DNA bending that is stabilized by the PU.1 domain may serve as an illustration of the hypothesis of DNA bending by phosphate neutralization. It has been demonstrated that when neutral methylphosphonates are introduced into DNA fragments, bending of the DNA occurs due to repulsion of the remaining anionic phosphates (Strauss and Maher, 1994). They proposed that binding of a cationic protein to DNA could have the same effect and it appears that PU.1 induces and stabilizes this type of bending in the DNA. There are seven sites of phosphate neutralization in the PU.1 DNA complex, on one face of the DNA helix.

Conclusions

The work accomplished in this project has been a significant contribution to our understanding of the way that ets proteins recognize DNA. We have successfully produced the first crystal structure of an ets protein and the model will serve as the basis to begin to describe the atomic detail for protein-DNA contacts. The contact regions have been evaluated by measurements in solution by NMR indicating structural features where molecular dynamics may contribute to DNA recognition. This study of the PU.1 molecule is the first study of an ets molecule where both crystal data and NMR solution data are produced.

Bibliography

Brown, L.F., Papadopoulos-Sergiou, A., Berse, B., Manseau, E.J., Tognazzi, K., Perruzzi, C.A., Dvorak, H.F. and Senger, D.R. (1994) Osteopontin expression and distribution in human carcinomas. *Am. J. Path.* **145:**610-623.

Clark, K.L., Halay, E.D., Lai, E., and Burley, S.K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364:**412-420.

Denhardt, D.T. and Guo, X. (1993) Osteopontin: A protein with diverse functions. *FASEB J.* **7:**1475-1482.

Donaldson, L.W., Petersen, J.M., Graves, B.J., and McIntosh, L.P. (1996) Solution structure of the ETS domain from murine Ets-1: A winged helix-turn-helix DNA binding motif. *EMBO J.* **15:**125-134.

Gardner, H.A.R., Berse, B., and Senger, D.R. (1994) Specific reduction in osteopontin synthesis by antisense RNA inhibits the tumorigenicity of transformed Rat1 fibroblasts. *Oncogene* **9:**2321-2326.

Guo, X., Zhang, P., Mitchell, D.A., Denhardt, D.T., and Chambers, A.F. (1995) Identification of a ras-activated enhancer in the mouse osteopontin promoter and its interaction with a putative ETS-related transcription factor whose activity correlates with the metastatic potential of the cell. *Mol. Cell. Biol.* **15:**476-487.

Hromas, R. and Klemsz, M. (1994) The ETS oncogene family in development, proliferation and neoplasia. *Inter. J. Hematol.* **59:**257-265.

Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C., and Maki, R.A. (1990) The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell* **61:**113-124.

Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A., and Ely, K.R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS domain-DNA complex. *Nature* **380:**456-460.

Liang, H., Mao, X., Olejniczak, E.T., Nettesheim, D.G., Yu, L., Meadows, R.P., Thompson, C.B., and Fesik, S.W. (1994) Solution structure of the ets domain of Fli-1 when bound to DNA. *Nature* **1:**871-876.

Moreau-Gachelin, F. (1994) Spi-1/PU.1: An oncogene of the Ets family. *Biochim. Biophy. Acta* **1198:**149-163.

Pio, F., Ni, C.-Z., Mitchell, R.S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R., and Ely, K.R. (1995) Co-crystallization of an ETS domain (PU.1) in complex with DNA: Engineering the length of both protein and oligonucleotide. *J. Biol. Chem.* **270:**24258-24263.

Pio, F., Kodandapani, R., Ni, C.-Z., Shepard, W., Klemsz, M., McKercher, S.R., Maki, R.A., and Ely, K.R. (1996) *J. Biol. Chem.* **271:**23329-23337.

Scott, D.K., Daniel, J.C., Xiong, X., Maki, R.A., Kabat, D., and Benz, C.C. (1994) Binding of an ets-related protein within the DNase-1 hypersensitive site of the HER2/neu promoter in human breast cancer cells. *J. Biol. Chem.* **269:**19848-19858.

Slamon, D., Godolphin, W., Jones, L., Holt, J., Wong, S., Keith, D., Levin, W., Stuart, S., Udove, J., Ullrich, A., and Press, M. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer cells. *Science* **244:**707-712.

Strauss, J.K., and Maher, L.J., III (1994) DNA bending by asymmetric phosphate neutralization. *Science* **266:**1829-1834.

Trimble, M.S., Xin, J.-H., Guy, C.T., Muller, W.J., and Hassell, J.A. (1993) PEA3 is overexpressed in mouse metastatic mammary adenocarcinomas. *Oncogene* **8:**3037-3042.

Wasylyk, B., Hahn, S.L., and Giovane, A. (1993) The Ets family of transcription factors. *Eur. J. Biochem.* **211:**7-18.

Welte, T., Garimorth, K., Philipp, S., Jennewein, P., Huck. C., Cato, A.C.B., and Doppler, W. (1994) Involvement of Ets-related proteins in hormone-independent mammary cell-specific gene expression. *Eur. J. Biochem.* **223:**997-1006.

Werner, M.H., Clore, G.M., Fisher, C.L., Fisher, R.J., Trinh, L., Shiloach, J., and Gronenborn, A.M. (1995), Erratum (1996) The solution structure of the human ETS-1DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **83:**761-771.

Werner, M.H., Clore, G.M., Fisher, C.L., Fisher, R.J., Trinh, L., Shiloach, J., and Gronenborn, A.M. (1996) Erratum to: The solution structure of the human ETS-1DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **83:**761-771.

Final Summary DAMD17-94-J-4439

Publications and Meeting Abstracts

Pio, F., Ni, C.-Z., Mitchell, R.S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R., and Ely, K.R. (1995) Co-crystallization of an ETS domain (PU.1) in complex with DNA: Engineering the length of both protein and oligonucleotide. *J. Biol. Chem.* **270:**24258-24263.

Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A., and Ely, K.R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS domain-DNA complex. *Nature* **380:**456-460.

Pio, F., Kodandapani, R., Ni, C.-Z., Shepard, W., Klemsz, M., McKercher, S.R., Maki, R.A., and Ely, K.R. (1996) *J. Biol. Chem.* **271:**23329-23337.

Pio, F., Kodandapani, R., Ni, C.-Z., Piccialli, G., McKercher, S., Klemsz, M., Maki, R.A., and Ely, K.R. Crystal Structure of PU.1 ETS Domain-DNA Complex: A New Pattern for Helix-Turn-Helix Recognition.: ASBMB/ASIP/AAI Joint Meeting, New Orleans, LA, 1996.

Ely, K.R., Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., McKercher, S., Klemsz, M., and Maki, R.A. Crystal Structure of PU.1 ETS Domain-DNA Complex: A New Pattern for Helix-Turn-Helix Recognition. International Union of Crystallography, Seattle, WA, 1996.

Personel

Ely, Kathryn R.	Principal Investigator
Assa-Munt, Nuria	Co-Investigator
Ni, Chao-Zhou	Crystallographer
Pio, Frédéric	Postdoctoral Associate
Morikis, Dimitri	NMR Spectroscopist
Light, James	Postdoctoral Associate
Ortiz, Maria	Postdoctoral Associate

APPENDIX

Two reprints appended.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 38, Issue of September 20, pp. 23329–23337, 1996
Printed in U.S.A.

New Insights on DNA Recognition by ets Proteins from the Crystal Structure of the PU.1 ETS Domain-DNA Complex*

(Received for publication, May 24, 1996, and in revised form, July 6, 1996)

Frédéric Pio‡§, Ramadurgam Kodandapani‡§, Chao-Zhou Ni‡, William Shepard¶, Michael Klemsz‖, Scott R. McKercher‡, Richard A. Maki‡, and Kathryn R. Ely‡ ‡‡**

*From the ‡La Jolla Cancer Research Center, The Burnham Institute, La Jolla, California 92037, ¶LURE, Université Paris-Sud, Bât. 209, 91405 Orsay Cedex, France, and **Neurocrine Biosciences, San Diego, California 92121*

Transcription factors belonging to the ets family regulate gene expression and share a conserved ETS DNA-binding domain that binds to the core sequence 5′-(C/A)GGA(A/T)-3′. The domain is similar to α+β ("winged") helix-turn-helix DNA-binding proteins. The crystal structure of the PU.1 ETS domain complexed to a 16-base pair oligonucleotide revealed a pattern for DNA recognition from a novel loop-helix-loop architecture (Kodandapani, R., Pio, F., Ni. C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R. A., and Ely, K. R. (1996) *Nature* 380, 456–460). Correlation of this model with mutational analyses and chemical shift data on other ets proteins confirms this complex as a paradigm for ets DNA recognition. The second helix in the helix-turn-helix motif lies deep in the major groove with specific contacts with bases in both strands in the core sequence made by conserved residues in α3. On either side of this helix, two loops contact the phosphate backbone. The DNA is bent (8°) but uniformly curved without distinct kinks. ETS domains bind DNA as a monomer yet make extensive DNA contacts over 30 Å. DNA bending likely results from phosphate neutralization of the phosphate backbone in the minor groove by both loops in the loop-helix-loop motif. Contacts from these loops stabilize DNA bending and may mediate specific base interactions by inducing a bend toward the protein.

Transcription factors bind to target DNA sequences to regulate metabolic functions such as growth and differentiation. Typically, the molecular scaffold for DNA recognition is conserved within a given family of DNA-binding proteins. In some cases the similarity of these scaffolds suggests an evolutionary relationship between different families or comparison of scaffolds reveals a structural similarity that was obscured by sequence comparisons alone.

A recently discovered family of regulatory proteins, the *ets* gene family, includes more than 45 members in a variety of organisms from *Drosophila* to humans (1, 2). These molecules play a role in normal development and have been implicated in malignant processes such as erythroid leukemia and Ewing's sarcoma. The DNA-binding domain of ets proteins is a conserved region (ETS domain) that is about 85 residues in length. Although ets proteins share a homologous sequence in the ETS domain, they differ in length and in the relative position of this domain. In some molecules, the ETS domain is found at the carboxyl terminus (*e.g.* PU.1 (3); ets-1 (4); ets-2 (5)), while in others the domain is located in the middle of the sequence (erg (6)), or in the amino-terminal region (elk-1 (7)). Flanking regions are thought to form other functional domains that influence protein-protein recognition or inhibitory domains that mask the DNA-binding site (8, 9).[1] In ets-1, an α-helix that is located in an inhibitory domain immediately NH_2-terminal to the ETS domain unfolds on DNA-binding (10). Regardless of the position of the ETS domain within the intact ets proteins, there is strong sequence homology in this conserved region.

We have determined the crystal structure of the ETS domain of the PU.1 transcription factor complexed to DNA (11). The domain is similar to α+β helix-turn-helix (HTH)[2] DNA-binding proteins and contacts a 10-base pair region of duplex DNA that is bent (8°) but uniformly curved without distinct kinks. The PU.1 domain assumes a tight globular structure with three α-helices and a four-stranded antiparallel β-sheet enclosing a hydrophobic core. The topology of the domain is similar to the structures of other ets family proteins fli-1 (12), murine ets-1 (13), and human ets-1 (14) determined in solution by NMR. The common molecular scaffold is similar to DNA-binding proteins such as CAP (15) and resembles "winged"-HTH proteins including HNF-3γ (16). ETS domains bind as a monomer to the core sequence 5′-(C/A)GGA(A/T)-3′.

The PU.1 domain contacts DNA from three sites: the recognition helix (α3) interacts with the GGAA core sequence in the major groove, while contacts with the phosphate backbone on either side of this site are made in the minor groove by two loops. Therefore, the PU.1 ETS domain binds DNA by a loop-helix-loop motif. One loop is formed between β-strands 3 and 4 (a "wing") and the other is a loop in the position of the turn in the HTH motif (α2-turn-α3). The protein-DNA contacts stabilize a uniform bending of the duplex DNA that likely is due to phosphate neutralization by the PU.1 domain. Surprisingly, the protein-DNA interactions reported in the NMR structure of a human ets-1-DNA complex (14) differed dramatically from this pattern, involving different contacts and significant DNA deformation. Because of this discrepancy, we chose to test the validity of the PU.1-DNA complex as a model for other ets proteins. As reported here, when the results of mutational analyses on a number of ets proteins are correlated with the structure of the PU.1-DNA complex and with chemical shift data measured with the fli-1 (12) and murine ets-1 (13) mole-

* This work was supported in part by USAMRDC Department of the Army Grants DMD17-94J-4439 (to K. R. E.), National Institute of Allergy and Infectious Disease Grant AI20194 (to R. A. M.), and National Cancer Institute Grant CA63489 (to K. R. E.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Contributed equally to the results presented in this study.

‖ Present address: Dept. of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN 46202.

‡‡ To whom correspondence should be addressed: The Burnham Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037. Tel.: 619-646-3135; Fax: 619-646-3196.

[1] M. Klemsz and R. A. Maki, unpublished results.

[2] The abbreviation used is: HTH, helix-turn-helix.

TABLE I
Crystallographic refinement statistics

R_{sym} (%)	3.3
Resolution range (Å)	6–2.1
Average B (Å^2)	31.65
Crystallographic R-factor (%)	22.5
R_{Free} (%)	28.7
Number of reflections used	22022 F $>2\sigma$(F)
Number of protein atoms	1486
Number of DNA atoms	1300
Number of solvent atoms	143

Root mean square deviation from ideal	r.m.s.	Target
Bond distance (Å)	0.012	(0.06)
Bond angles (degrees)	1.629	(10)
Dihedral angles (degrees)	1.575	(20)

cules, the loop-helix-loop scaffold is confirmed as a general model for DNA recognition by ets proteins. This pattern defines a new class of HTH DNA-binding proteins. The molecular pattern of DNA recognition by ets proteins is compared to other HTH proteins for which crystal structures of the protein-DNA complexes are available.

EXPERIMENTAL PROCEDURES

PU.1 DNA Complex—A recombinant fragment encompassing residues 160–272 from the murine ets protein PU.1 was crystallized in complex with a 16-base pair oligonucleotide representing a consensus PU.1 DNA-binding site (3) as described previously (17). The complex crystallized in space group $C2$ with a = 89.1, b = 101.9, c = 55.6 Å, and β = 111.2°. There are two complexes in the asymmetric unit. The length of the oligonucleotide was critical for crystallization and the oligonucleotide used to form the complex permitted end-to-end stacking of the DNA in the crystal lattice with the formation of pseudo-base pairing by the overhanging A and T bases.

Crystallographic Analyses—The initial structure analysis of the complex solved by the MIRAS method was reported (11). For this first phase of the study, a native data set and four heavy atom data sets were collected using a Rigaku RU200 rotating anode x-ray source and two San Diego Multiwire Systems area detectors. The initial data sets were collected from flash frozen crystals at 2.3-Å resolution. To refine the structure further, another native data set extending to 2.1 Å was collected at the LURE synchrotron source in Orsay, France. Diffraction data were collected at station D41 interfaced with the Mark III multiwire proportional area detector. Data sets were processed using MOSFLM (18) and ROTAVATA, AGROVATA, and TRUNCATE in the CCP4 package (19). In the present study, this native data set was scaled to the data collected in the home laboratory by Wilson scaling and the synchrotron data were incorporated into the refinement. The programs PHASES (20), FRODO (21), and X-PLOR (22) were used for structure solution, model building, and refinement. The current R-factor is 22.5 for 6 to 2.1 Å data (22,022 reflections). The average overall B-factor for 2929 non-hydrogen atoms (1486 protein atoms + 1300 DNA atoms + 143 solvent oxygens) is 31.6 Å^2. The refinement statistics are presented in Table I. There were 11 disordered residues at the amino terminus of the domain and 14 disordered residues at the carboxyl terminus of the recombinant fragment that were excluded from the model. These residues were not ordered even when the resolution was extended to 2.1 Å. For all residues representing the complete ETS domain (residues 171–258), the electron density was clear and permitted unambiguous fitting of both backbone and side chain atoms. More solvent atoms have been added to the model. Only minimal changes in the configuration of some side chains were evident in the high resolution map. The stereochemistry of all main chain torsion angles in the domain fall within energetically favorable limits (Fig. 1) indicating that no segment of the domain is denatured or randomly configured. The DNA was clearly defined even in the first MIRAS map.

Analyses of DNA Helical Parameters—To analyze the stereochemical basis for the uniform bending observed in the oligonucleotide bound in complex to PU.1, the DNA superstructure was measured (23, 24) and four parameters were calculated that describe the conformation of the DNA bases and the phosphate backbone. The values were calculated (excluding the 5′ A overhang) to analyze helical parameters along the length of the oligonucleotide and to compare these with standard B-DNA parameters. The geometry of dinucleotide steps was analyzed for three rotational angles defining twist, tilt, or roll and for one translational distance, *i.e.* rise. The values for these parameters are presented in Table II.

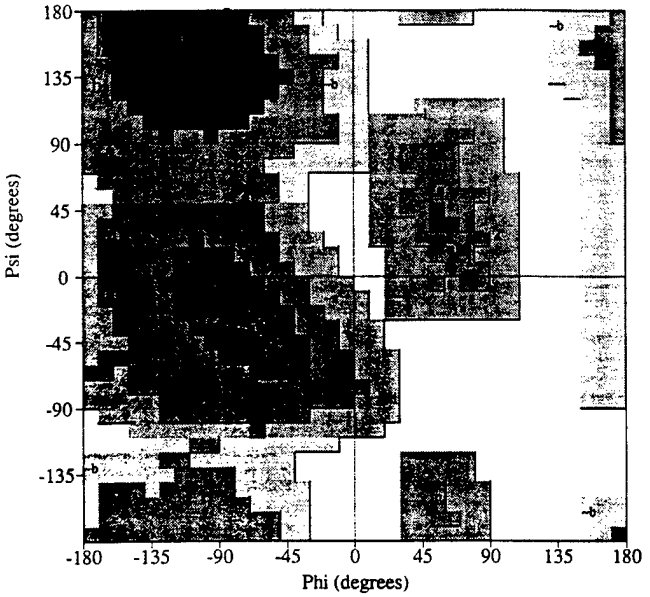

FIG. 1. **Ramachandran diagram for the current model of the PU.1 ETS domain.** This diagram presenting $\phi\psi$ angles (46) was produced using the PROCHECK programs (47). Glycine residues are represented by *triangles*. Various regions of the plot with different levels of shading are indicated with the *darkest shaded areas* corresponding to the energetically most favorable $\phi\psi$ angles.

Sequence Alignments and Structural Comparisons—Sequence alignments for ets proteins were made using GENEWORKS. The individual sequences were collected from the SWISSPROT data base and regions corresponding to the ETS domains were excised from the full-length protein before the alignment process began (25). The results of this comparison are presented in Fig. 2. Sequence comparisons between members of different families of HTH proteins were made using the program QUANTA (Molecular Simulations, Inc.) especially when structure-based alignments were utilized. To search structure data bases to identify proteins with similar overall scaffolds to the PU.1 domain, the algorithm DALI developed by Chris Sander (26) was used. For structural comparisons of HTH proteins, coordinates were obtained from the Brookhaven Protein Data Bank (27): 434 cro repressor (code 3CRO), λ repressor (code 1LMB), CAP (code 1CGP), and heat shock factor (code 2HTS). The coordinates for HNF-3γ were kindly provided by Dr. S. Burley. The actual structural comparisons/graphical analyses were performed using Quanta (Molecular Simulations, Inc.) and the Alberta/Caltech program TOM based on FRODO (21).

RESULTS AND DISCUSSION

The similarity of the structural organization of the ETS domains of PU.1 (11), fli-1 (12), and ets-1 (13, 14) and the presence of a conserved hydrophobic core suggests that this overall scaffold will be highly conserved in all members of the family. To facilitate comparisons, the sequences of the ETS domains of 33 members of the ets family are aligned (Fig. 2). The sequences of this domain in a number of ets proteins are identical for two or more species, representing a significant level of homology within the family. The results of mutational substitutions in a number of ets proteins are tabulated in Table III.

Hydrophobic Core—The importance of the hydrophobic core was verified by site-directed mutagenesis of the PU.1 domain (11). Of the 14 strictly conserved residues in the domain, seven are found in the hydrophobic core. Single substitution of glycine for five of these residues in PU.1 (Fig. 3) resulted in loss of DNA binding. Two of these core residues also contact the DNA phosphate backbone. The peptide amide nitrogen of Leu^{174} interacts with O2P of C-22 and the side chain NE-1 of Trp^{215} forms a hydrogen bond with O1P from T-23. Mutation of tryp-

TABLE II

DNA helical parameters of the 16-base pair oligonucleotide bound to PU.1

DNA structural parameters were refined in X-PLOR (22) and then analyzed using the programs developed by Babcock and Olson (24). For comparison, typical twist angles for B-DNA are 34.3°, roll angles are 0°, and rise values are 3.38 Å.

	Base pair	Inter-base pair			Slide (Å)	Intra-base pair	
		Helical twist (°)	Roll (°)	Rise (Å)		Propeller twist (°)	Buckle (°)
1	A-T	36.09	−0.06	3.18	0.13	−18.22	11.26
2	A-T	39.67	−0.48	3.23	0.06	−16.18	13.88
3	A-T	34.10	−6.19	3.30	−0.71	−14.38	1.39
4	A-T	36.09	0.76	3.20	−1.00	−17.25	3.38
5	G-C	27.50	6.59	3.57	−0.75	−8.02	2.87
6	G-C	29.06	7.29	3.11	0.24	2.10	−17.28
7	G-C	33.02	3.96	3.47	0.59	7.77	5.90
8	G-C	27.44	6.75	3.21	−0.37	−14.98	7.43
9	A-T	36.02	9.00	3.10	0.16	−21.87	10.59
10	A-T	39.21	3.24	3.35	−0.61	−19.29	5.58
11	G-C	24.41	−0.84	3.38	−0.94	−13.13	−10.09
12	T-A	37.07	3.75	3.29	0.98	−12.30	−7.49
13	G-C	32.70	9.93	3.30	−0.14	−6.40	3.66
14	G-C	33.29	4.99	3.23	−0.02	−10.53	−8.09
15	G-C					−9.96	−2.26

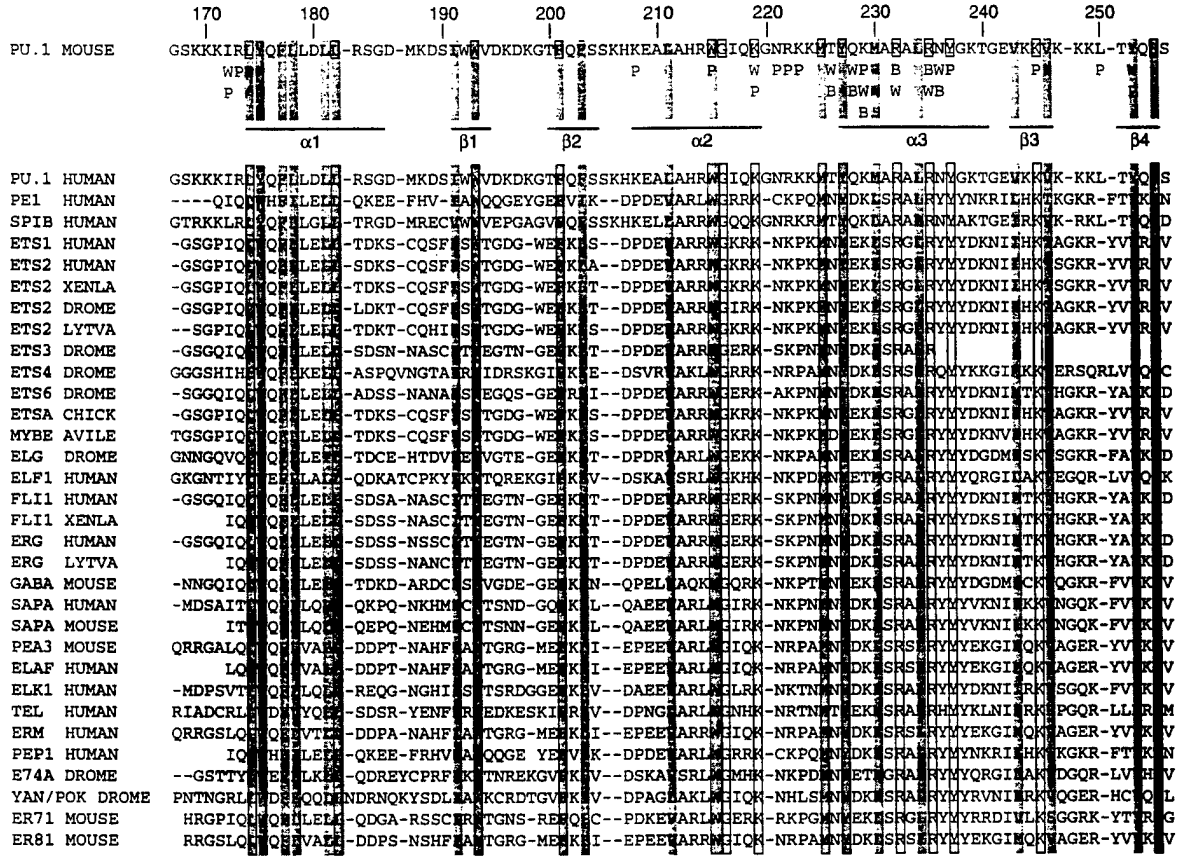

FIG. 2. **Sequence alignment of the DNA-binding domain of 33 members of the ets family.** The amino acid sequence of PU.1 is listed at the top of the figure and residues that are strictly conserved in the family are enclosed in *boxes*. The sequences were obtained from the SWISSPROT data base and original citations for the sequences are given in the data base. Secondary structural features of the PU.1 ETS domain are indicated above the alignment. Directly under the PU.1 sequence, the residues that contact DNA are indicated: *B*, base interaction; *P*, phosphate backbone interaction; *W*, water-mediated interaction. Residues found in the hydrophobic core in PU.1 and expected to be located in the hydrophobic interior of all ets proteins are *shaded*. In some cases, the sequences for ets proteins for two or several species are identical, and therefore only one sequence has been listed to avoid duplication.

tophan 215 to arginine results in loss of DNA binding in ets-1 (28, 29; see Table III). Substitutions in the hydrophobic core affect DNA binding probably because the changes disrupt the tight globular structure of the domain. Residues 174 and 215 are doubly critical for DNA binding since they represent both important structural residues in the domain core and actual DNA contact residues. In summary, residues in the hydrophobic core are critical for the formation of the overall scaffold for ets recognition.

Molecular Scaffold of ETS Domains—To evaluate the conservation of this scaffold within the ets family, the α-carbon backbones of PU.1 (11) and fli-1 (12) domains were superimposed utilizing both sequence homology and secondary structure similarities. For this purpose, a single model from the ensemble of structures deposited in the data bank was used for the NMR-derived fli-1 structure. This scaffold provides the

TABLE III

Mutations in the DNA-binding domain of ets family proteins that abolish DNA binding

The reference for each mutational substitution is given in parentheses with the protein studied.

ets Protein	Residue in PU.1[a]	Single mutation	Multiple mutations
PU.1 (11)	174H, D	L → G	
PU.1 (11)	178H	L → G	
ets-1 (29)	Multiple		174H, D, 175H, 177H, 178H
ets-1 (28)	185	K → P	
ets-1 (28)	191H	I → T	
PU.1 (11)	193H	W → G	
ets-1 (28)	194	T → I	
ets-1 (28)	196	D → G	
ets-1 (28)	201H	F → L	
PU.1 (11)	201H	F → G	
PU.1 (11)	203H	F → G	
ets-1 (28)	212	A → V	
ets-1 (28)	214	R → G	
ets1 (28, 29)	215H, D	W → R	
PU.1 (11)	215H, D	W → G	
ets-1 (28)	219D	K → *X*[b]	
PU.1 (11)	219D	K → G	
ets-1 (28)	222D	K → *X*[b]	
ets-1 (28)	227H	Y → C	
fli-1 (12)	228D	D → H, Q, K	
ets-1 (28)	232D	R → *X*[b]	
fli-1 (12)	232D	R → D, K, N	
PU.1 (11)	232D	R → G	
fli-1 (29)	234H	L → V	
ets-1 (29)	Multiple		234H, 235, 236, 237
fli-1 (12)	235D	R → K, D, N, E	
PU.1 (11)	235D	R → G	
fli-1 (12)	236D	Y → V	
ets-1 (48)	242	I → E, G, P, V	
ets-1 (28)	243H	I → T	
PU.1 (11)	245D	K → G	
ets-1 (28)	248	K → I	
ets-1 (28)	254H	F → L	

[a] Residue numbers of the PU.1 sequence are given to facilitate direct comparison with the sequence alignment in Fig. 2; H indicates a residue in the hydrophobic core of the PU.1 domain and D indicates residues which contact DNA in the PU.1-DNA complex, either directly or by water-mediated interactions.

[b] *X*, substitution by any amino acid.

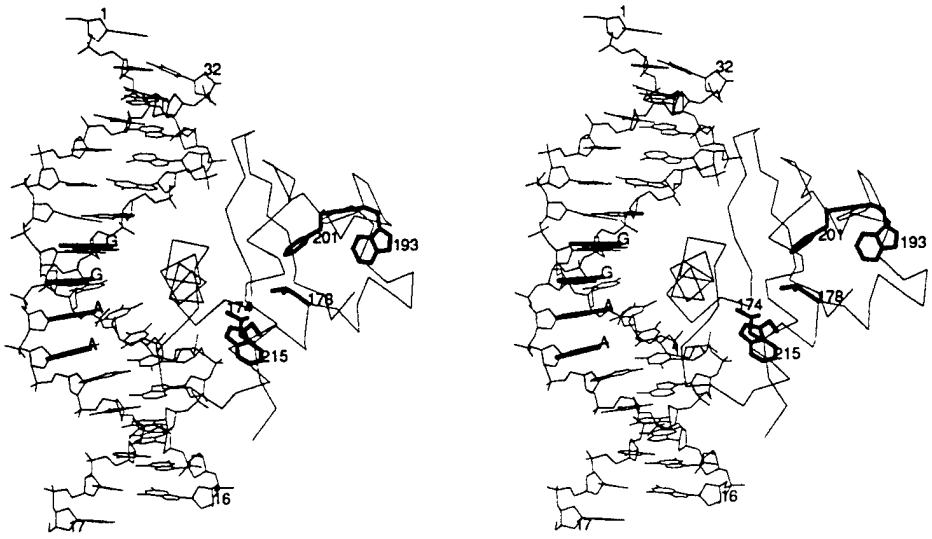

FIG. 3. **Stereodiagram of the PU.1-ETS domain DNA complex.** The α-carbon backbone for residues 171–258 is shown bound to DNA with the bases in the GGAA core in *bold lines*. The ETS module is composed of three α-helices and a four-stranded antiparallel β-sheet enclosing a hydrophobic core. There are seven strictly conserved residues in this core (Fig. 2). Substitution of glycine for each of the five core residues in PU.1, shown on the model, abolishes DNA binding.

framework for the three structural features arranged in a loop-helix-loop pattern that mediate precise DNA binding by the PU.1 domain. In order to delineate the loop-helix-loop motif in other ets domains and to predict whether this motif is the paradigm for ets recognition, we also superimposed the α-carbon skeleton of the fli-1 domain onto the PU.1 backbone bound to the DNA (Fig. 4). Since this is one of an ensemble of structures from the NMR study, detailed comparisons are not possible. However, general comparisons are useful to establish overall structural similarities between the two related molecules. Although the structure of the fli-1-DNA complex was not determined, it should be noted that the published structure of the fli-1 domain (12) reflects a bound conformation since the NMR experiments were conducted on a 98-residue protein frag-

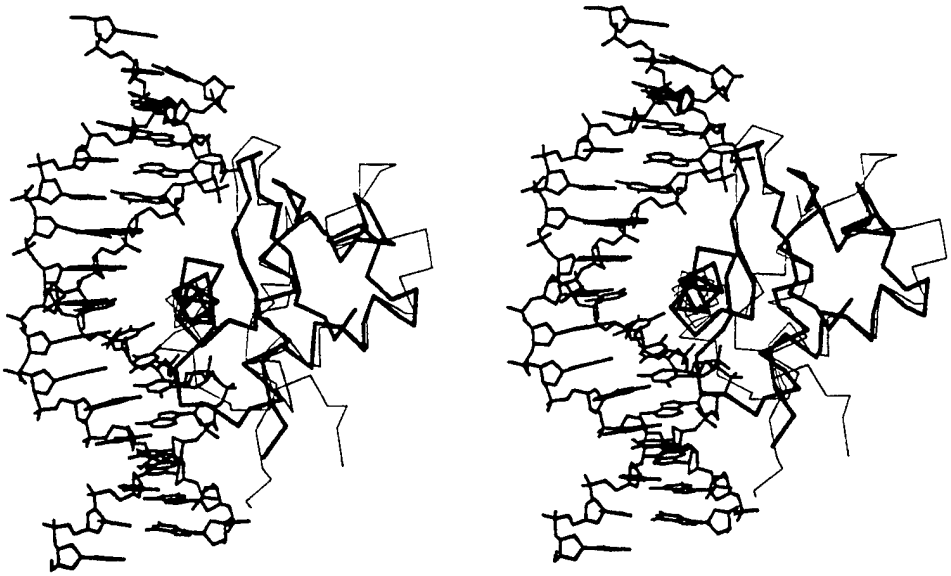

FIG. 4. **Comparison of PU.1 and fli-1 ETS domains.** In this stereo image, the α-carbon backbone of the fli-1 ETS domain (*thin line*; residues 276–373), determined in solution by NMR (12) was superimposed on the PU.1 backbone (*bold line*) using constraints to match similar structural features. The DNA shown in the figure is the oligonucleotide bound to the PU.1 domain.

ment complexed to a 16-base pair oligonucleotide.

As shown in Fig. 4, there is close similarity in the overall scaffold of the ETS domains but several other features of the superposition are worth noting. First, the positions of the four conserved residues that contact DNA are very similar in PU.1 and fli-1. In PU.1, two conserved arginines, 232 and 235, make hydrogen bonds with the bases GGA of the PU core sequence. Arg^{235}(NH-2) forms a hydrogen bond with G-8(O-6) while Arg^{232}(NH-1) makes hydrogen bonds with two bases G-9(O-6) and A-10(N-6) on one strand and a water-mediated contact with T-23(O-4) on the opposite strand. These arginines are strictly conserved in all members of the ets family and the GGA sequence is the consensus DNA sequence recognized by the ets proteins. Therefore, these interactions are expected to be reproduced in all ets protein-DNA complexes. When the fli-1 domain is superimposed on PU.1, the side chains of conserved arginines 232 and 235 in the recognition helix are within hydrogen-bonding distance of the same bases in the GGAA core sequence in the major groove. Substitution of these residues by any other amino acid, even closely related hydrophilic amino acids results in loss of DNA recognition in PU.1, fli-1, and other ets proteins (see Table III). Conserved lysines, residues 219 in the loop (HTH) and 245 in the wing contact the phosphate backbone in PU.1 and are in a position to make the same contacts in fli-1. Mutational substitutions for Lys^{219} in PU.1 (11) and the equivalents of Lys^{219} and Lys^{222} (see Table III) in fli-1 (12) or ets-1 (28) disrupt DNA binding, presumably due to the loss of the phosphate backbone interactions. In fli-1, the equivalents of Lys^{222} and Met^{225} in PU.1 (from the HTH loop) and residues 248/249 (from the wing loop) were identified within 4 Å of DNA by intermolecular NOEs (12). Chemical mapping experiments with the murine ets-1 molecule suggested a similar pattern with a major groove contact zone and interactions with both adjacent minor grooves (30).

DNA Conformation in the PU.1 ETS Domain-DNA Complex—The PU.1 ETS domain contacts DNA over a 10-base pair area. The DNA is bent by 8° in the complex but does not deviate significantly from B-form DNA (see Table II). As can be seen in Fig. 4, the DNA is uniformly curved over the length of the 16-base pair fragment. There is an average helical twist of 33°, with 10.8 base pairs per turn and an average rise per base pair of 3.2 Å. The minor groove is slightly enlarged (~2–3 Å from the mean) in the GGAA region at the midpoint of the oligonucleotide. A "spine" of water molecules, similar to that observed in the crystal structure of a B-DNA dodecamer (31), is located in the minor groove from bases 8 to 12. Binding of the ETS domain induces a DNase I-hypersensitive site 3′ to the C-26 base in the core sequence (30). This site is probably exposed on the face of the DNA opposite to where the protein binds as a result of the expansion of the minor groove (Fig. 3).

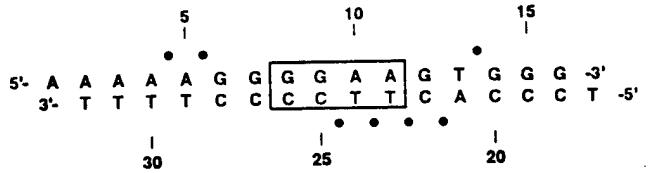

FIG. 5. **Sequence of the oligonucleotide bound to the PU.1 protein in the crystal structure.** The GGAA recognition core sequence as well as the bases on the complementary strand are enclosed in a *box*. The PU.1 domain makes contacts with bases on both strands within this core. The *dots* designate seven phosphates that are neutralized by interactions with basic residues. With the exception of the phosphate at base 14, all of these phosphates lie on one face of the DNA helix.

The DNA bending that is stabilized by the PU.1 domain may serve as an illustration of the hypothesis of DNA bending by phosphate neutralization. It has been demonstrated, by the introduction of neutral methylphosphonate analogues in DNA fragments bearing polyadenylate tracts (32) that bending of the DNA occurs when the phosphate charges are neutralized on one face of the DNA helix, due to repulsion of the remaining anionic phosphates. It was proposed (32) that binding of proteins with cationic surfaces to DNA could also cause the DNA double helix to "spontaneously relax" toward the surface where cationic amino acids neutralized phosphate anions through formation of salt bridges. The PU.1 ETS domain makes neutralizing contacts with phosphate groups on one face of the DNA helix, involving consecutive phosphates on either side of the major groove. The sites of phosphate neutralization are shown on the DNA sequence in Fig. 5. On the GGAA strand, neutralizing contacts with the phosphate backbone 5′ to the core sequence are made by Lys^{208} and Lys^{245} from the wing. On the complementary strand, the phosphate contacts are 5′ to the core sequence as well as with the phosphate backbone within the core: Arg^{173}, Lys^{219}, and Lys^{223} from the HTH loop and Lys^{229} from helix α3. As predicted by the neutralization exper-

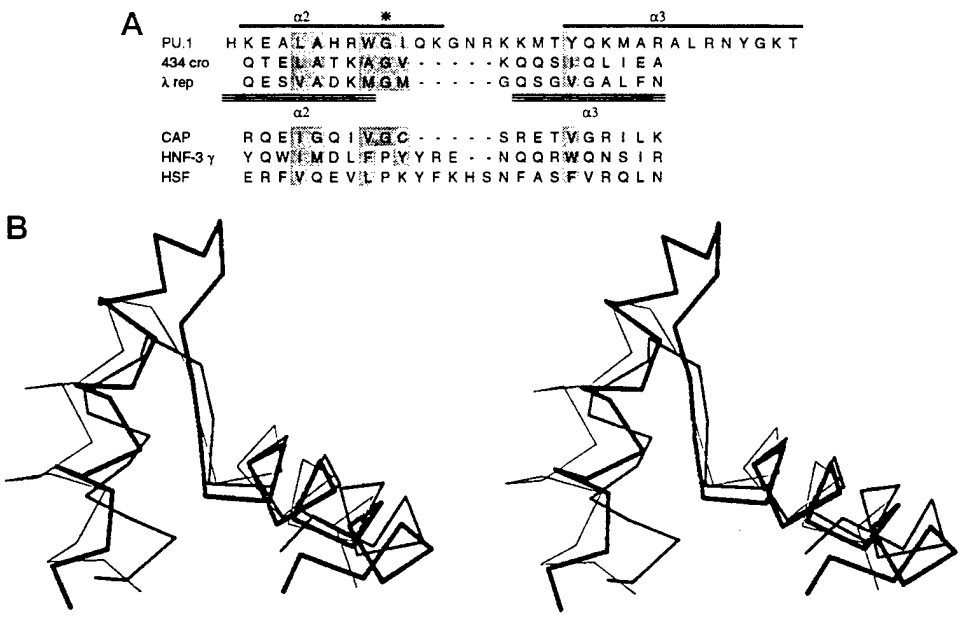

FIG. 6. **Comparison of the HTH motif in PU.1 with other proteins in the HTH superfamily.** *Panel A*, sequence alignment of residues that form the HTH motif in PU.1 with classic HTH proteins 434 cro repressor (42) and λ repressor (41), and with α+β-HTH proteins CAP (15), HNF-3γ (16), and heat shock factor (*HSF*) (40). The structural helices of the PU.1 ETS domain are indicated by *solid bars* above the PU.1 sequence and the helices in the bacterial repressors are indicated by *open bars* below the λ repressor sequence. This figure was adapted from Fig. 1 in Ref. 43. Residues that are *shaded* represent positions in classic HTH that are generally hydrophobic or small (Gly or Ala) in these proteins. The glycine that is conserved in the bacterial HTH proteins is marked with an *asterisk*. Note that helix α2 in PU.1 is one turn longer than the counterpart in the bacterial proteins, yet when the HTH motifs of the repressors are superimposed on the PU.1 HTH, the glycine in the last turn of the PU.1 α2 helix is equivalent to the conserved glycine in the turn of the bacterial proteins (not shown). *Panel B*, the HTH motifs of PU.1 (*thick line*), CAP (medium line; Ref. 15), and heat shock factor (*thin line*; Ref. 40) are superimposed for comparison. The α3 recognition helix is on the *right* in the photograph. Note that the relative orientation of the two helices is closely similar in the three molecules, but the configuration of the residues in the turn between the helices is different. The turn in the PU.1 domain is seven residues in length which is intermediate between the extremes reported for the family of HTH proteins (43, 44).

iments (32), the cationic surface of the PU.1 domain binds to the DNA causing a bend of the duplex oligonucleotide toward the ETS module that is within the range (~10°) of curvature estimated experimentally. The bend is toward the "neutral surface," *i.e.* toward the protein. Two of these phosphate interactions in the minor groove involve conserved residues, Lys^{219} from the HTH loop and Lys^{245} from the wing. Thus the loop-helix-loop pattern may influence both DNA recognition and DNA bending.

This type of charge neutralization is not seen in all protein-induced DNA bends. For example, the TATA-binding protein binds with extensive phosphate backbone interactions to the TATA element (33). Yet in this case the DNA is sharply kinked away from the protein contacts. In CAP (15) salt bridges and other hydrogen bonds to phosphate groups stabilize a severely kinked DNA conformation with DNA bent at 90°.

Interactions with the phosphate backbone are seen in numerous DNA-binding proteins, but these contacts are often hydrogen bonds and not salt bridges. The hypothesis (32) states that neutralization of charge by lysines and arginines results in excess repulsive electrostatic forces that can maintain bending of the DNA double helix (34). The moderate DNA bending seen in the complexes of oligonucleotides with paired homeodomains (35, 36) or HNF-3γ (16) may also result from phosphate neutralization, since these proteins form phosphate-side chain salt bridges with 4 or 3 arginines, respectively. However, the neutralizing contacts are not as extensive as those seen in the PU.1-DNA complex.

The complementarity of the loop-helix-loop motif of fli-1 with the DNA from the PU.1 complex also suggests that, like PU.1, other ETS domains may not significantly deform DNA from B-DNA conformation but to date there is not much biochemical data in the literature on DNA bending by ETS domains. In one study of the ETS domain from the Elk-93 protein, circular permutation analyses indicated that DNA binding by the Elk-93 fragment did not induce significant bending of DNA (37). In contrast, in the human ets-1-DNA complex (14), the DNA was kinked at a 60° angle due to intercalation of a tryptophan side chain. The equivalent of this tryptophan, tyrosine 175 in PU.1, is found in the hydrophobic core and is not in position to intercalate. Substitution of glycine for this tyrosine in PU.1 does not affect DNA binding (11). In fli-1 (12), the equivalent tryptophan is buried in the hydrophobic core and was not listed among residues in close proximity (≤4Å) to DNA. Thus, the molecular basis for kinked DNA cannot be understood in the context of contacts seen in the PU.1-DNA complex (11) or inferred in the fli-1 complex (12). DNA bending by phosphate neutralization is not apparent in the ets-1-DNA complex, since only one lysine and one arginine form phosphate-side chain salt bridges. The arginine is the equivalent of Arg^{235} in PU.1 that forms a hydrogen bond with base G-8 in the GGA core.

Target Specificity—The superimposed models in Fig. 4 suggest that a loop-helix-loop scaffold that brings together conserved amino acids and conserved DNA bases is a general mode of DNA recognition by ets proteins. Yet, ets transcription factors bind to the GGA(A/T) core motif in the context of specific promoters. To begin to identify residues that influence target specificity, it is necessary to look for mutations of non-conserved residues that affect DNA binding. Of the 14 absolutely conserved residues in the domain, seven contact DNA in the PU.1 complex. These contacts would be expected to be maintained for all ets-DNA complexes. In studies of a number of members of the ets family, mutations have been reported that affect DNA binding. These mutations, summarized in Table III, can now be correlated with the atomic model of the PU.1-DNA complex. Some of these residues are conserved residues, but others are unique to a particular molecule.

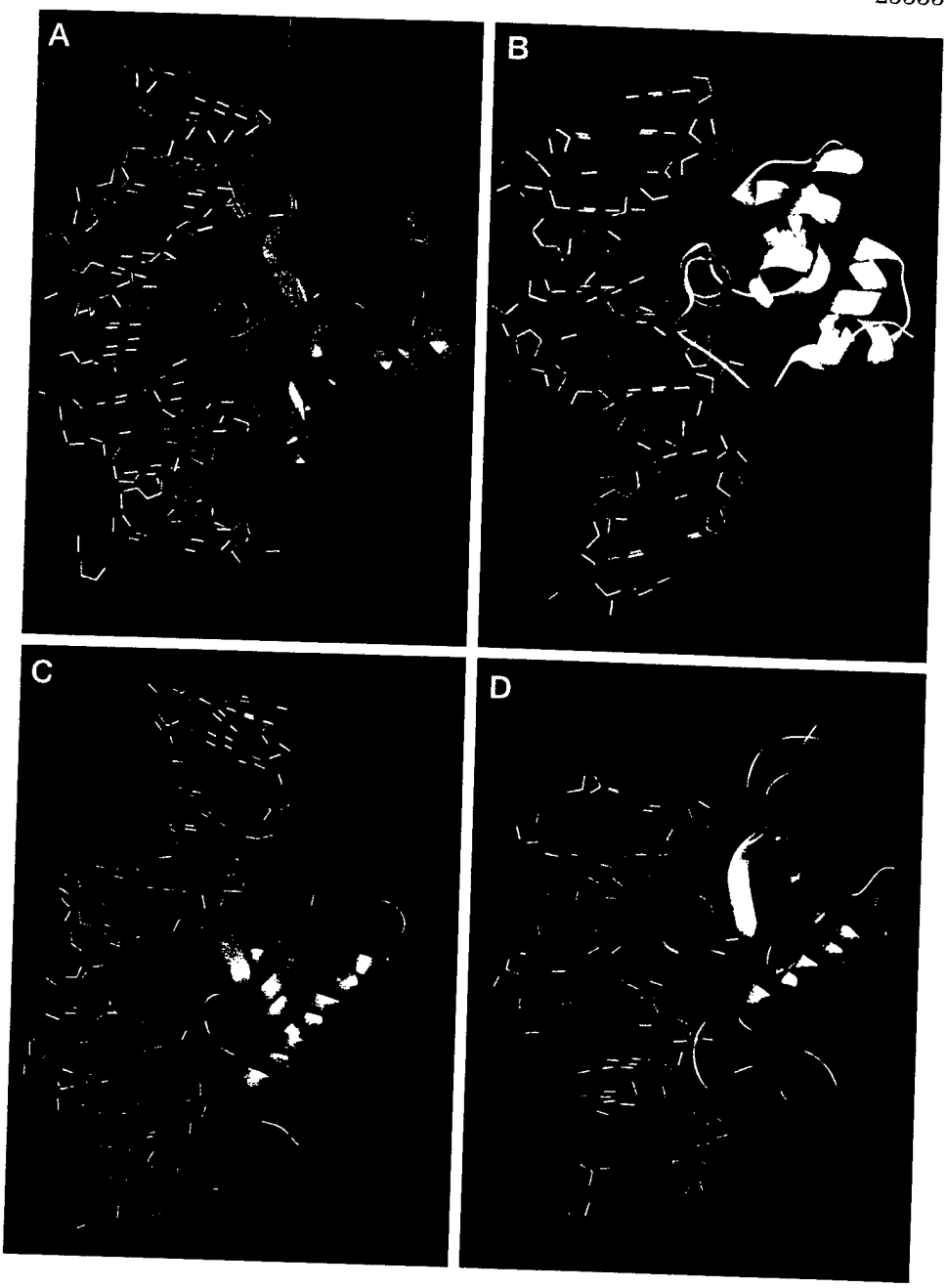

FIG. 7. **Comparison of protein-DNA complexes in HTH proteins.** The loop-helix-loop pattern of DNA recognition in the PU.1 complex (*panel A*) is compared to a classic HTH protein 434 cro repressor (42) (*panel B*), to an $\alpha+\beta$ HTH protein CAP (15) (*panel C*), and to a winged-HTH protein HNF-3γ (16) (*panel D*). In each of these complexes, the recognition helix makes contact in the major groove. The contacts of the PU.1 domain with DNA are more extensive and include interactions from two loops in the minor grooves on either side of the major groove where recognition helix α3 binds.

It should be emphasized that PU.1 contacts both strands at the GGAA core. Interactions are made by conserved residues as well as residues where sequence variability exists in the ets family. Therefore, ets recognition requires specific base contacts with the GGAA sequence and the bases on the complementary strand. For example, it has been shown that a single residue converts DNA recognition of ets proteins from GGAA to GGAT. When a lysine in chicken ets-1 (equivalent to residue 229 in PU.1) is altered to threonine found in this position in Elf-1 and E74, the resultant protein exhibits a restricted selectivity for GGAA like the Elf-1/E74 proteins and the reverse mutation causes the converse change in DNA recognition (38). In the PU.1 complex, Lys^{229} is located in the recognition helix and makes a water-mediated contact to base C-25 on the antisense strand at the GGAA core. There is a water network located in the major groove at the GGAA site. Twelve well defined water molecules are hydrogen-bonded to the bases and also form a hydrogen-bonded network between the two strands. This water network may contribute to the stability of the duplex and consequently influence specific DNA recognition. Since the side chain of lysine is long, it is possible that the contact of a shorter residue such as threonine would not bind to this water network and could contact a different base, *i.e.* T-23. The water network itself could also change. Or, the interchange of lysine↔threonine could permit DNA contact reflecting the stereochemical difference in size of adenine *versus* thymine bases.

HTH Motif—All of the direct contacts with specific bases in the PU.1-DNA complex are made by residues in the α3 recognition helix. Two non-conserved residues, Thr^{226} and Gln^{228}, at the amino-terminal end of this helix, make water-mediated contacts with bases C-25 and C-26, respectively, that are base paired to guanines 8 and 9 in the core GGAA sequence. Both of these residues are unique to PU.1/SpiB in the ets family, so these may represent PU.1-specific contacts.

Tyr^{227}, which is strictly conserved in the ets family, is located in the hydrophobic interior of the protein. While the phenyl ring of this tyrosine is buried, the hydroxyl group is exposed and lies within 3.6 Å of G-6(O1P). This residue was not included in our list of DNA contacts using a conservative cut-off

of 3.2 Å for hydrogen bonds/ionic interactions. Although this interaction may not occur in PU.1, with a simple side chain rotation, a hydrogen bond is possible with the phosphate backbone. This may be an example of a contact made by a conserved residue that influences DNA recognition by selected family members. Substitution of cysteine for this tyrosine abolishes DNA binding in ets-1 (28).

In Fig. 6A, the sequence of the HTH motif of PU.1 is compared with the sequence of "classic" bacterial HTH proteins and other winged-HTH proteins. The glycine required in the turn between helices in HTH proteins (39) is also conserved in this position in ETS domains, although the α2 helix is one turn longer than the helix in HTH proteins. In PU.1, the glycine lies in the last turn of this helix. This glycine and other hydrophobic residues in α2 and α3 stabilize the arrangement of these two helices in HTH proteins. Even this pattern of conserved hydrophobic residues is seen in ets proteins. In other winged-HTH proteins, HNF-3γ (16) or heat shock factor (40), the sequence similarities are not as apparent. These two proteins have prolines in the equivalent position of the conserved glycine and the presence of this proline may influence the configuration in the "turn." On the other hand, ets proteins may exhibit a helical arrangement that is structurally closer to that in "classic" HTH proteins. When HTH elements of PU.1 and HTH molecules such as λ (41) or 434 cro (42) repressors are superimposed, the glycine is in a structurally equivalent position (not shown). Moreover, the overall pattern of docking of the recognition helix in the major groove is quite similar when 434 cro repressor (42), CAP (15), and PU.1 are compared bound to DNA (Fig. 7). The major difference is the fact that the recognition helix in PU.1 docks deep in the major groove with contacts to the bases involving residues along the entire length of the helix, while DNA contacts in CAP and other classic HTH proteins are made from residues at the amino-terminal portion of the helix.

None of the related proteins in the HTH superfamily actually contact DNA by residues in the HTH turn (43, 44). This novel DNA contact may be possible in PU.1, as well as other ets proteins, because the connecting segment between helices is more of a loop than a turn. The corresponding HTH motifs of heat shock factor (40) and CAP (15) are compared to PU.1 in Fig. 6*b*. But it is not simply the length of the "turn" or "loop" in the HTH motif that accounts for this DNA contact in PU.1, since other eukaryotic HTH proteins contain even longer connecting segments (43, 44) and yet do not contact DNA by this structural feature, for example HNF-3γ (16). Thus the contacts made by this loop in PU.1 illustrate a new DNA contact that, to date, is unique to the ets proteins as the newest members of the HTH superfamily.

Loops and Minor Groove Contacts—Since the sequences in the HTH loop as well as the loop (wing) between strands β3 and β4 are not strictly conserved among members of the ets family, these residues may be important sites for specific recognition by individual members of the family. In the PU.1-DNA complex, these two loops contact the minor groove through interactions with the phosphate backbone closest to the major groove. It is also interesting to note that the length of both of the contact loops differs among members of the family, with the PU.1 loop containing an "extra" glycine at residue 220 and lacking a glycine after residue 247. Other residues in these loops may also provide specific contacts to bases in other ets proteins. For example, the change of arginine→aspartic acid (equivalent to 244 in PU.1) affects DNA binding in Elk-1 (45).

Since ets proteins bind DNA as monomers, it could be expected that there would be extensive contacts to stabilize the interaction. HNF-3γ also binds DNA as a monomer (16). In the HNF-3γ complex, three regions were involved in DNA recognition: the recognition helix and two wings. The location of the first wing between the last two strands in the β-sheet corresponds topologically to the wing in PU.1, but contacts from the second wing emanate from a loop at the COOH terminus of the domain. The structural equivalent of this second loop is absent in PU.1. In CAP, the major DNA contacts are made from the recognition helix. This protein binds DNA as a dimer. The surface area on CAP that is buried on DNA binding is 1187 $Å^2$. Similarly, the surface area buried when 434 cro repressor binds DNA is 1306 $Å^2$. But with the formation of the DNA complex with the PU.1 ETS domain, 1701 $Å^2$ surface area is buried. The significantly greater surface area of the PU.1 domain covered reflects the extensive protein-DNA contact region extending for more than 30 Å (11).

The PU.1-DNA model suggests that residues from the two loops contribute the critical interactions for recognition of bases other than the conserved GGAA core when the core is embedded in specific promoter sequences. The loops approach segments of the DNA that are adjacent to the conserved core sequence and therefore these interfaces are stereochemically suitable to permit sequence-specific interactions by a given family member while maintaining the consensus interactions at GGA(A/T). Moreover, the contacts from these loops may mediate specific base interactions by stabilizing a bend toward the protein. Future extensive mutational studies of amino acids that contact DNA are needed to identify these residues. Ultimately, crystal structures of other ets proteins complexed to DNA can be compared to distinguish unique DNA contacts.

Acknowledgments—We thank Dr. Roger Fourme and the staff at the LURE synchrotron for consultation and support for this project. In addition, we are grateful to John Knight and Rick Mitchell for synthesis and purification of oligonucleotides, and to Kelly Riddle-Hilde for preparing the manuscript for publication.

REFERENCES

1. Moreau-Gachelin, F. (1994) *Biochim. Biophys. Acta* **1198,** 149–163
2. Wasylyk, B., Hahn, S. L., and Giovane, A. (1993) *Eur. J. Biochem.* **211,** 7–18
3. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., and Maki, R. A. (1990) *Cell* **61,** 113–124
4. Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J., and Papas, T. S. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 7294–7298
5. Reddy, E. S., and Rao, V. N. (1988) *Oncogene Res.* **3,** 239–246
6. Reddy, E. S., Rao, V. N., and Papas, T. S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 6131–6135
7. Rao, V. N., Huebner, K., Isobe, M., ar-Rushdi, A., Croce, C. M., and Reddy, E. S. (1989) *Science* **244,** 66–70
8. Wasylyk, C., Kerchaert, J.-P., and Wasylyk, B. (1992) *Genes Dev.* **6,** 965–974
9. Lim, F., Kraut, N., Frampton, J., and Graf, T. (1992) *EMBO J.* **11,** 643–652
10. Petersen, J. M., Skalicky, J. J., Donaldson, L. W., McIntosh, L. P., Alber, T., and Graves, B. J. (1995) *Science* **269,** 1866–1869
11. Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R. A., and Ely, K. R. (1996) *Nature* **380,** 456–460
12. Liang, H., Mao, X., Olejniczak, E. T., Nettesheim, D. G., Yu, L., Meadows, R. P., Thompson, C. B., and Fesik, S. W. (1994) *Nature Struct. Biol.* **1,** 871–876
13. Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1996) *EMBO J.* **15,** 125–134
14. Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J., and Gronenborn, A. M. (1995) *Cell* **83,** 761–771
15. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science* **253,** 1001–1007
16. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature* **364,** 412–420
17. Pio, F., Ni, C.-Z., Mitchell, R. S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R. A., and Ely, K. R. (1995) *J. Biol. Chem.* **270,** 24258–24263
18. Leslie, A. (1994) *MOSFLM User Guide*, Mosflm version 5.20, MRC Laboratory of Molecular Biology, Cambridge, England
19. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr.* **50,** 760–764
20. Furey, W., and Swaminathan, S. (1990) *Am. Cryst. Assoc. Meeting* **18,** 73
21. Jones, T. A. (1985) *Methods Enzymol.* **115,** 157–171
22. Brünger, A. T. (1992) *X-PLOR Manual*, Version 3.1, Yale University, New Haven, CT
23. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T., and Berman, H. M. (1996) *Acta Crystallogr.* **D52,** 57–64
24. Babcock, M. S., and Olson, W. K. (1994) *J. Mol. Biol.* **237,** 98–124
25. Holm, L., and Sander, C. (1994) *Nucleic Acids Res.* **22,** 3600–3609
26. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
27. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. Jr., Brice, M. D.,

Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542
28. Mavrothalassitis, G., Fisher, R. J., Smyth, F., Watson, D. K., and Papas, T. S. (1994) *Oncogene* **9,** 425–435
29. Wang, C.-Y., Petryniak, B., Ho, I.-C., Thompson, C. B., and Leiden, J. M. (1992) *J. Exp. Med.* **175,** 1391–1399
30. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., and Graves, B. J. (1992) *Genes Dev.* **6,** 975–990
31. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol.* **151,** 535–556
32. Strauss, J. K., and Maher, L. J., III (1994) *Science* **266,** 1829–1834
33. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) *Nature* **365,** 520–527
34. Crothers, D. M. (1994) *Science* **266,** 18–19
35. Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell* **80,** 639–650
36. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell* **82,** 709–719
37. Shore, P., Bisset, L., Lakey, J., Waltho, J. P., Virden, R., and Sharrocks, A. D. (1995) *J. Biol. Chem.* **270,** 5805–5811
38. Bosselut, R., Levin, J., Adjadj, E., and Ghysdael, J. (1993) *Nucleic Acids Res.* **21,** 5184–5191
39. Brennan, R. G., and Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 1903–1906
40. Harrison, C. J., Bohm, A. A., and Nelson, H. C. M. (1994) *Science* **263,** 224–227
41. Beamer, L. J., and Pabo, C. O. (1992) *J. Mol. Biol.* **227,** 177–196
42. Mondragon, A., and Harrison, S. C. (1991) *J. Mol. Biol.* **219,** 321–334
43. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053–1095
44. Brennan, R. G. (1992) *Curr. Opin. Struc. Biol.* **2,** 100–108
45. Janknecht, R., Zinck, R., Ernst, W. H., and Nordheim, A. (1994) *Oncogene* **9,** 1273–1278
46. Ramachandran, G. N., and Sasiekharan, V. (1968) *Adv. Protein Chem.* **23,** 283–437
47. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26,** 283–291
48. Soudant, N., Albagli, O., Dhordain, P., Flourens, A., Stéhelin, D., and Leprince, D. (1994) *Nucleic Acids Res.* **22,** 3871–3879

A new pattern for helix–turn–helix recognition revealed by the PU.1 ETS-domain–DNA complex

Ramadurgam Kodandapani*, Frédéric Pio*, Chao-Zhou Ni*, Gennaro Picciallì†, Michael Klemsz‡, Scott McKercher*, Richard A. Maki*§ & Kathryn R. Ely*

* La Jolla Cancer Research Center at The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA
† Dipartimento di Chimica Organica e Biologica, Universita degli Studi di Napoli Federico II, 80134 Napoli, Italy
§ Neurocrine Biosciences, San Diego, California 92121, USA

THE Ets family of transcription factors, of which there are now about 35 members[1,2], regulate gene expression during growth and development. They share a conserved domain of around 85 amino acids[3] which binds as a monomer to the DNA sequence 5′-C/AGGAA/T-3′. We have determined the crystal structure of an ETS domain complexed with DNA, at 2.3-Å resolution. The domain is similar to α + β (winged) 'helix–turn–helix' proteins and interacts with a ten-base-pair region of duplex DNA which takes up a uniform curve of 8°. The domain contacts the DNA by a novel loop–helix–loop architecture. Four of the amino acids that directly interact with the DNA are highly conserved: two arginines from the recognition helix lying in the major groove, one lysine from the 'wing' that binds upstream of the core GGAA sequence, and another lysine, from the 'turn' of the 'helix–turn–helix' motif, which binds downstream and on the opposite strand.

The PU.1 *[Spi-1, Spfi-1]* transcription factor is an Ets protein expressed in haematopoietic cells[4,5]. PU.1 is a regulatory protein for differentiation of monocytes and macrophages and for B-cell maturation (reviewed in ref. 2). The ETS domain of PU.1 was co-crystallized with a 16 base-pair oligonucleotide containing the recognition sequence[6]. The structure was solved by the multiple isomorphous replacement and anomalous scattering (MIRAS) method (Table 1). The electron density was clearly defined (Fig. 1) for residues 171 to 258, which encompasses the entire conserved ETS domain. The PU.1 domain assumes a tight globular structure ($33 \times 34 \times 38$ Å^3) formed by three α-helices and a four-stranded antiparallel β-sheet (Fig. 1). The domain topology is similar to the structures of other Ets family proteins Fli-1 (ref. 7), murine Ets-1 (ref. 8) and human Ets-1 (ref. 9) determined in solution by NMR. The structural studies revealed a common folding pattern for ETS domains that is similar to α + β helix–turn–helix (HTH) DNA-binding proteins including CAP[10] and resembles 'winged' HTH proteins such as GH5 (ref. 11), HNF-3γ (ref. 12) and HSF (ref. 13). There are three sites of protein–DNA contact: the recognition helix (α3), the loop between β-strands 3 and 4 (a 'wing') and the turn in the HTH motif (α2–turn–α3). The turn between α2 and α3 is longer than the equivalent in many other HTH proteins, and is actually a loop. The DNA-binding motif in PU.1, and probably other members of the Ets family, can be described more appropriately as a loop–helix–loop motif. Therefore the large Ets family defines a new variant subclass of the helix–turn–helix DNA-binding proteins with a novel mode of DNA recognition.

The protein–DNA contacts in the PU.1 complex are detailed in Fig. 2. Four strictly conserved residues on the surface of the domain are likely to be important for DNA binding by all members of the Ets family. Arg 232 and Arg 235 emanate from helix α3 and contact bases in the GGAA sequence in the major groove. These contacts represent the core structure for DNA

‡ Present address: Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, Indiana 46202-5120, USA.

recognition by members of the Ets family because they involve both strictly conserved amino acids and bases in the consensus sequence recognized by these transcription factors (see Fig. 3*b*). The equivalent arginines 81 and 84 in Ets-1 (ref. 9) do not contact the GGAA bases, but intermolecular nuclear Overhauser effects between these arginines and DNA were observed in the Fli-1 NMR studies[7]. Lys 245 extends from β3 just adjacent to the loop ('wing'), and Lys 219 is located in the 'loop' of the HTH motif. Lys 245 contacts the phosphate backbone of the GGAA strand in the minor groove upstream from the core sequence (Fig. 3*c*) and Lys 219 forms a salt bridge with the phosphate backbone of the opposite strand downstream of the GGAA core (Fig. 3*d*). Substitutions of glycine at each of these four conserved sites abolished DNA binding, confirming the functional importance of these contacts (see Fig. 2).

Mutations of conserved residues that contact the phosphate backbone also affect DNA binding. Substitution of glycine at Leu 174 or Trp 215 abolished DNA binding in PU.1. Similarly, substitution of any amino acid in Ets-1 (ref. 14) at the equivalent of PU.1 residues Lys 219 and Arg 222 that bind the phosphate backbone disrupted DNA binding. These minor-groove contacts might represent a conserved pattern for protein 'docking' in the Ets family. In Fli-1 (ref. 7), the equivalents of Leu 174, Lys 219 and Lys 222 showed large chemical shifts on DNA binding in the NMR studies (the counterpart of Trp 215 was buried).

Water molecules also participate in protein–DNA recognition in the PU.1 complex (Fig. 2). There are 27 well-ordered solvent molecules around the DNA. Solvent molecules in the major groove are hydrogen-bonded to the bases and also form a hydrogen-bonded network between the two strands that might contribute to the stability of the duplex and consequently influence specific DNA recognition. Conserved Arg 232 and Arg 235 each form direct and water-mediated contacts with the bases. Three other residues also contact DNA bases through water molecules: Thr 226, Gln 228 and Asn 236. These residues are not conserved in the Ets family and might represent interactions that are unique to the PU.1 protein. Thr 226 and Gln 228, at the amino-terminal end of helix α3, make water-mediated contacts with bases C25 and C26 respectively that are base-paired to guanines 8 and 9 in the core sequence.

TABLE 1 Structure determination and refinement

	Native	Hg	*I* (29)	*I* (13)	*I* (31)
Phasing statistics					
Resolution (Å)	2.3	3.0	2.9	3.0	2.8
Observed reflections	60,095	25,081	20,709	20,512	23,308
Unique reflections	20,105	14,902	13,258	12,910	15,397
Completeness (%)	97	79	65	69	68
R_{sym} (%)*	5.0	3.6	4.0	4.3	3.6
R_{iso} (%) to 3.0 Å†		13.0	14.4	15.9	13.0
Number of sites		2	2	2	2
For isomorphous data ($I/\sigma \geq 3$)					
Phasing power‡		1.33	1.76	1.04	0.98
To resolution (Å)		3.0	3.0	3.0	3.0
R_{Cullis}§		0.62	0.57	0.68	0.67
For anomalous data ($I/\sigma \geq 3$)					
Phasing power		1.0	1.41	1.13	1.43
To resolution (Å)		3.0	3.0	3.0	3.0
Mean figure of merit (10–3.0 Å) is 0.65.					
Refinement statistics					
Resolution range		8–2.3 Å			
Average *B* (Å^2)		20.1			
Crystallographic *R*-factor (%)		23.7			
R_{free} (%)[16]		29.9			
Number of reflections used		16,898 $F > 3\sigma(F)$			
Number of protein atoms		1,486			
Number of DNA atoms		1,300			
Number of solvent atoms		88			

The crystallization of the PU.1 ETS domain (residues 160–272) with a 16-bp synthetic DNA oligonucleotide containing the recognition sequence was described previously[6]. Crystals formed in the space group C_2 with $a = 89.1$, $b = 101.9$, $c = 55.6$ Å and $\beta = 111.2°$, with two complexes in the asymmetric unit. **Phase determination**. Four heavy-atom derivatives were prepared by soaking crystals of the native complex and by co-crystallizing iodinated oligonucleotides with the PU.1 domain. The locations of the iodinated bases are indicated in Fig. 2. Multiple isomorphous replacement phases, including anomalous data, were calculated. The package PHASES[17] was used to refine heavy-atom positions, *B*-factor/occupancies and to calculate phases to 3.0-Å resolution with an overall figure of merit of 0.65. The initial MIRAS map (3.0 Å) was improved by solvent flattening by the method of Wang[18] and with non-crystallographic density averaging. **Model building and refinement**. The improved MIRAS electron-density map was used to build the model with the interactive graphics programs TOM based on FRODO[19] and O[20]. The density for the DNA helix was a prominent feature of the map. To fit the DNA, an 'ideal' B–DNA duplex was generated with the program QUANTA (Molecular Simulations, Inc.) and fitted to the density as a rigid body. After the DNA was positioned, a polyalanine chain was constructed with the BONES option of the Alberta/Caltech program TOM. Subsequently side chains for all residues with clear electron density were added to the model. There were 11 disordered residues at the N terminus of the domain and 14 disordered residues at the C terminus so these amino acids were not included in the model. For all other residues representing the complete ETS domain, the electron density was clear (see Fig. 1) and allowed unambiguous fitting of both backbone and side-chain atoms. Manual adjustments of individual DNA bases were made to fit the electron density. In the program X-PLOR[21], the stereochemistry of the protein was optimized to bond and angle parameters developed by Engh and Huber[22] and for DNA by using parameters of Parkinson *et al.*[23]. Weak restraints were placed on all ribose conformations. One cycle of simulated annealing at 3,000 K (ref. 24) was followed by cycles of manual model building, positional refinement and *B*-factor refinement. More data were added as the refinement progressed in increments: 3, 2.8, 2.6 and 2.3 Å. A total of 88 solvent oxygens ($\langle B \rangle = 22$ Å^2) have been added to the model at this stage of the refinement. Main-chain torsion angles for all non-glycine residues fall within energetically favourable Ramachandran boundaries[25]. The r.m.s. difference for 84 α-carbon atoms in the two complexes in the asymmetric unit is 0.35 Å.

* R_{sym} is $\sum |I - \langle I \rangle| / \sum \langle I \rangle$.

† R_{iso} is $\sum ||F_{PH}| - |F_P|| / \sum |F_P|$, where $|F_P|$ and $|F_{PH}|$ are structure factors for the protein and derivative, respectively.

‡ Phasing power is the r.m.s. value of $|F_H|/E$, where E is residual lack of closure.

§ $R_{Cullis} = \sum ||F_{PH}| \pm |F_P| - |F_{H(calc)}|| / \sum |F_{PH} - F_P|$ for centric reflections, where $F_{H(calc)}$ is the calculated heavy-atom structure factor.

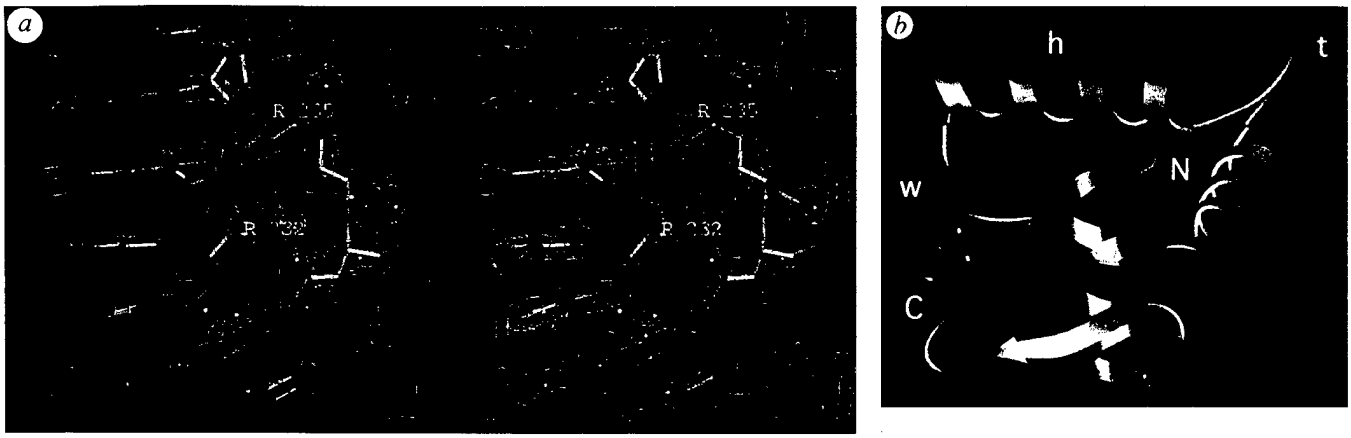

FIG. 1 Overall structure of the PU.1–DNA complex. *a*, Stereoview of the refined 2.3-Å (1.5σ) $2F_o - |F_c|$ electron-density map at the protein–DNA interface. Interactions of protein (gold), DNA (red) and water (white) in the major groove at the GGAA core recognition sequence are shown. The two strictly conserved residues, Arg 232 and Arg 235 from recognition helix α3, make direct contact with bases in the major groove. There is a tight network of water molecules at this site in the major groove (Fig. 2). *b*, Ribbon drawing of the PU.1 ETS domain. The module (residues 171–258) is composed of three α-helices and a four-stranded antiparallel β-sheet. In the interior of the domain, a hydrophobic core is formed with 19 side chains including seven strictly and eight highly conserved residues. The major structural features that contact the DNA are indicated: the recognition helix α3 (h), the turn in the HTH motif (t) and the loop between β-strands 3 and 4 (w) corresponding to the 'wing' in these proteins. At the N-terminal end of the fragment, helix α1 begins at residue 172. The C-terminal segment, which is disordered in the PU.1- DNA complex, assumes an α-helical conformation in the unbound Ets-1 NMR structure[4]. This segment might unfold in PU.1 with DNA binding. *c*, Space-filling model of the PU.1 ETS domain–DNA complex. Protein–DNA interactions include both major and minor groove contacts over a distance of 30 Å. The PU.1 transcription factor (gold) binds to DNA as a monomer, so it is not surprising that extensive DNA contact sites exist in addition to the recognition sequence to stabilize binding. HNF-3γ (ref. 12) and GH5 (ref. 11) also bind to target DNA as monomers. In the HNF-3γ–DNA complex, three regions were involved in DNA recognition: the recognition helix and two 'wings'. The location of the first 'wing' between the last two strands in the β-sheet corresponds topologically to the 'wing' in PU.1, but contacts from the second 'wing' emanate from a loop at the C terminus of the domain. The structural equivalent to the second 'wing' is absent in PU.1.

FIG. 3 PU.1 DNA complex. *a*, The 16-bp oligonucleotide bound in complex to the PU.1 ETS domain is shown in grey, with the GGAA sequence coloured red. The ETS domain is represented by an orange ribbon model with the side chains for four conserved residues that contact DNA shown. When glycine was introduced at each of these sites, DNA binding was lost (Fig. 2). *b*, Detailed close-up view, showing that Arg 232 and Arg 235 from the recognition helix make hydrogen bonds with the bases GGA of the PU core sequence. Arg 235(NH2) forms a hydrogen bond with G8(O6), whereas Arg 232(NH1) makes hydrogen bonds with two bases G9(O6) and A10(N6). These arginines are strictly conserved in all members of the Ets family, and the GGA sequence is the consensus DNA sequence recognized by the Ets proteins. Therefore the interactions shown here represent the paradigm for Ets recognition, which is expected to be reproduced in all Ets protein–DNA complexes. *c*, Interaction of the 'wing': Lys 245(NZ) contacts the phosphate backbone at G6(O2P). *d*, Interaction of Lys 219(NZ) from the loop in the HTH motif, which contacts the phosphate backbone at C22(O3P) and T23(O2P). This figure was generated with the graphics program QUANTA (Molecular Simulations, Inc.).

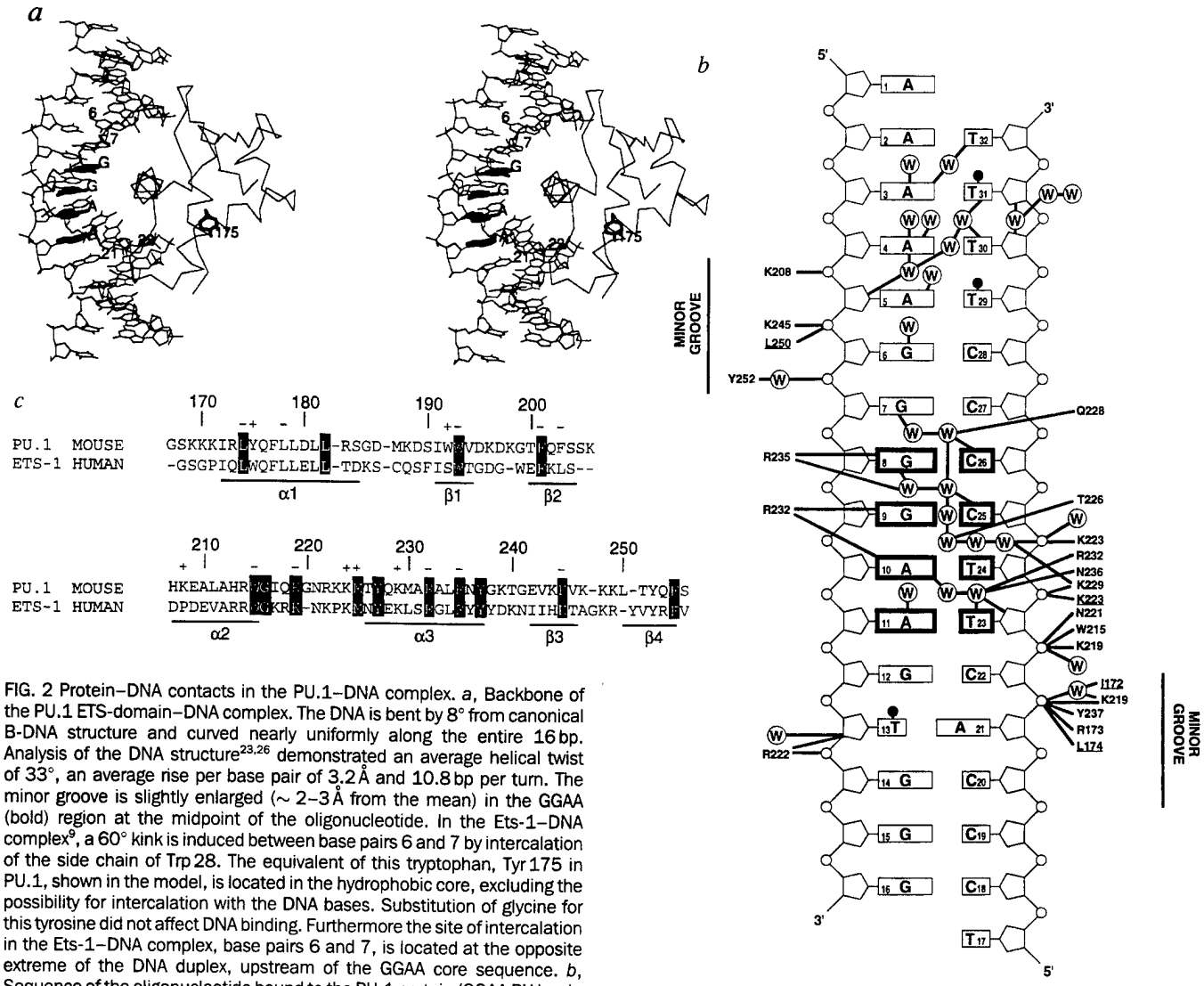

FIG. 2 Protein–DNA contacts in the PU.1–DNA complex. *a*, Backbone of the PU.1 ETS-domain–DNA complex. The DNA is bent by 8° from canonical B-DNA structure and curved nearly uniformly along the entire 16 bp. Analysis of the DNA structure[23,26] demonstrated an average helical twist of 33°, an average rise per base pair of 3.2 Å and 10.8 bp per turn. The minor groove is slightly enlarged (~ 2–3 Å from the mean) in the GGAA (bold) region at the midpoint of the oligonucleotide. In the Ets-1–DNA complex[9], a 60° kink is induced between base pairs 6 and 7 by intercalation of the side chain of Trp 28. The equivalent of this tryptophan, Tyr 175 in PU.1, shown in the model, is located in the hydrophobic core, excluding the possibility for intercalation with the DNA bases. Substitution of glycine for this tyrosine did not affect DNA binding. Furthermore the site of intercalation in the Ets-1–DNA complex, base pairs 6 and 7, is located at the opposite extreme of the DNA duplex, upstream of the GGAA core sequence. *b*, Sequence of the oligonucleotide bound to the PU.1 protein (GGAA PU box in bold lines). Residues that contact the DNA through main-chain atoms are underlined. Well-defined solvent molecules located within 3.2 Å of protein or DNA atoms are identified by an encircled W. Contacts from residues of the 'wing' are made with the nucleotides upstream of the GGAA sequence, and residues from the loop in the HTH motif interact with the opposite strand, downstream of the GGAA site. The direction of the DNA was confirmed by the location of the three iodinated bases (13,29,31; black dots) used for phase calculation. Seven of the residues that contact DNA are strictly conserved and four others are highly conserved. *c*, Sequence alignment of the PU.1 and Ets-1 ETS domains, representing extremes of evolutionary divergence in the family. Residues strictly conserved in all Ets proteins are shown in black boxes; dashes indicate gaps within the family. Numbering and secondary structural features of the PU.1 domain are indicated. The results of mutational analysis when glycine was substituted for a residue are also shown. The effects of the interchanges are labelled + or − above the sequence, indicating that DNA binding was retained or abolished. Mutations were generated essentially as described[27].

The turn in the HTH motif is actually a loop, and because the sequences in this loop as well as the loop ('wing') between strands β3 and β4 are not strictly conserved among members of the Ets family, these residues might be important sites for specific recognition by individual members of the family. In fact, the lengths of both of the contact loops differ among members of the family, with the PU.1 loop containing an 'extra' glycine at residue 220 and lacking a glycine after residue 247. Such conformational differences are expected between family members, but the contrast between the PU.1 and Ets-1 complexes was unexpected. The striking distinction in the mode of DNA contact by the PU.1 and Ets-1 domain could reflect extreme evolutionary divergence between members of the Ets family. Alternatively, it should be noted that the Ets-1–DNA complex was formed under denaturing conditions[9,15] and it is possible that the Trp intercalation occurred early during the renaturation step with subsequent protein refolding.

Future extensive mutational studies of amino acids that contact DNA in Ets proteins are needed to identify residues that mediate recognition of a specific DNA sequence by a given family member. Ultimately, crystal structures of other Ets proteins complexed to DNA must be compared to distinguish unique DNA contacts. □

Received 12 January; accepted 19 February 1996.

1. Wasylyk, B., Hahn, S. L. & Glovane, A. *Eur. J. Biochem.* **211,** 7–18 (1993).
2. Moreau-Gachelin, F. *Biochim. biophys. Acta* **1198,** 149–163 (1994).
3. Karim, F. *et al. Genes Dev.* **4,** 1451–1453 (1990).
4. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C. & Maki, R. A. *Cell* **61,** 113–124 (1990).
5. Moreau-Gachelin, F., Mattei, M. G., Tambourin, R. & Tavitian, A. *Oncogene* **4,** 1449–1456 (1989).
6. Pio, F. *et al. J. biol. Chem.* **270,** 24258–24263 (1995).
7. Liang, H. *et al. Nature struct. Biol.* **1,** 871–875 (1994).
8. Donaldson, L. W., Petersen, J. M., Graves, B. J. & McIntosh, L. P. *EMBO J.* **15,** 125–134 (1996).
9. Werner, M. H. *et al. Cell* **83,** 761–771 (1995).
10. Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253,** 1001–1007 (1991).

11. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. *Nature* **362,** 219–223 (1993).
12. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. *Nature* **364,** 412–420 (1993).
13. Harrison, C. J., Bohm, A. A. & Nelson, H. C. M. *Science* **263,** 224–227 (1994).
14. Mavrothalassitis, G., Fisher, R. J., Smyth, F., Watson, D. K. & Papas, T. S. *Oncogene* **9,** 425–435 (1994).
15. Werner, M. H., Clore, G. M., Gronenborn, A. M., Kondoh, A. & Fisher, R. J. *FEBS Lett.* **345,** 125–130 (1994).
16. Brünger, A. T. *Nature* **355,** 472–475 (1992).
17. Furey, W. & Swaminathan, S. *Am. crystallogr. Ass. Meeting* **18,** 73 (1990).
18. Wang, B. C. *Methods Enzymol.* **115,** 90–112 (1985).
19. Jones, T. A. *Methods Enzymol.* **115,** 157–171 (1985).
20. Jones, T. A., Zhou, J., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47,** 110–119 (1992).
21. Brünger, A. T. *X-PLOR Manual* Version 3.1 (Yale Univ. Press. New Haven, CT. 1992).
22. Engh, R. A. & Huber, R. *Acta crystallogr.* **A57,** 392–400 (1991).
23. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. *Acta crystallogr.* **D52,** 57–64 (1996).
24. Brünger, A. T., Krokowski, A. & Erickson, J. W. *Acta crystallogr.* **A46,** 585–593 (1990).
25. Ramachandran, G. N. & Sasiekharan, V. *Adv. Protein Chem.* **23,** 283–438 (1968).
26. Babcock, M. S. & Olson, W. K. *J. molec. Biol.* **237,** 98–124 (1994).
27. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. *Gene* **77,** 51–59 (1989).

ACKNOWLEDGEMENTS. This work was supported by grants from the U.S. Army and NIH. R.K. and F.P. made equal contributions to this study. We thank J. Knight and R. Mitchell for synthesis and purification of DNA oligonucleotides, M. Hasham for excellent graphics illustrations, and K. Riddle-Hilde for preparing the manuscript for publication.

CORRESPONDENCE AND MATERIALS. Requests to be addressed to K.R.E.